

Designing Adaptive Instruction for Teams: a Meta-Analysis

Robert A. Sottolare¹  · C. Shawn Burke² ·
Eduardo Salas³ · Anne M. Sinatra¹ ·
Joan H. Johnston¹ · Stephen B. Gilbert⁴

© The Author(s) 2017. This article is an open access publication

Abstract The goal of this research was the development of a practical architecture for the computer-based tutoring of teams. This article examines the relationship of team behaviors as antecedents to successful team performance and learning during adaptive instruction guided by Intelligent Tutoring Systems (ITSs). Adaptive instruction is a training or educational experience tailored by artificially-intelligent, computer-based tutors with the goal of optimizing learner outcomes (e.g., knowledge and skill acquisition, performance, enhanced retention, accelerated learning, or transfer of skills from instructional environments to work environments). The core contribution of this research was the identification of behavioral markers associated with the antecedents of

✉ Robert A. Sottolare
robert.a.sottolare.civ@mail.mil

C. Shawn Burke
sburke@ist.ucf.edu

Eduardo Salas
eduardo.salas@rice.edu

Anne M. Sinatra
anne.m.sinatra.civ@mail.mil

Joan H. Johnston
joan.h.johnston.civ@mail.mil

Stephen B. Gilbert
gilbert@iastate.edu

¹ US Army Research Laboratory, Orlando, FL 32826, USA

² University of Central Florida, Orlando, FL 32816, USA

³ Rice University, Houston, TX 77005, USA

⁴ Iowa State University, Ames, IA 50011, USA

team performance and learning thus enabling the development and refinement of teamwork models in ITS architectures. Teamwork focuses on the coordination, cooperation, and communication among individuals to achieve a shared goal. For ITSs to optimally tailor team instruction, tutors must have key insights about both the team and the learners on that team. To aid the modeling of teams, we examined the literature to evaluate the relationship of teamwork behaviors (e.g., communication, cooperation, coordination, cognition, leadership/coaching, and conflict) with team outcomes (learning, performance, satisfaction, and viability) as part of a large-scale meta-analysis of the ITS, team training, and team performance literature. While ITSs have been used infrequently to instruct teams, the goal of this meta-analysis make team tutoring more ubiquitous by: identifying significant relationships between team behaviors and effective performance and learning outcomes; developing instructional guidelines for team tutoring based on these relationships; and applying these team tutoring guidelines to the Generalized Intelligent Framework for Tutoring (GIFT), an open source architecture for authoring, delivering, managing, and evaluating adaptive instructional tools and methods. In doing this, we have designed a domain-independent framework for the adaptive instruction of teams.

Keywords Team tutoring · Teamwork · Taskwork · Adaptive instruction · Team-based instruction · Team learning · Team performance · Team satisfaction · Team viability · Collective · Collective training · Collaborative learning

Introduction

While one-to-one human tutoring has claimed to be significantly more effective than one-to-many instructional methods (e.g., traditional classroom instruction; Bloom 1984; VanLehn 2011), it is neither a practical nor affordable solution in large organizations (e.g., academic, corporate, or military; Sottolare and Proctor 2012). The use of computer-based tutoring programs for learning has seen a renewed interest in training and educational domains and one-to-one computer-based tutoring continues to emerge as a practical alternative to one-to-one human tutoring.

One-to-one tutoring via Intelligent Tutoring Systems (ITSs) provides tailored experiences to engage individual learners, offers an effective means to enhance their learning and performance, but have focused mainly on well-defined educational domains (e.g., cognitive tasks involving problem solving or decision-making). Tutors for physics, mathematics, and software programming make up the bulk of the ITSs produced today. A recent review of artificial intelligence in education (AIED) meta-analyses by du Boulay (2016) noted investigations by VanLehn (2011), Ma et al. (2014), Kulik and Fletcher (2015), Steenbergen-Hu and Cooper (2013, 2014), and Pane et al. (2014). Each meta-analysis provided a range of results for effectiveness in the context of one-to-one tutoring in individual instructional domains.

In recent years, military trainers (U.S. Army Training, and Doctrine Command, TRADOC 2011; North Atlantic Treaty Organization 2012) have been requesting ITSs capabilities that can support training and education of both individuals *and* teams. Teams, the basic building blocks of military organizations, are important to demonstrating progress toward goals, developing solutions, and meeting challenges associated

with organizational missions. As there are only a few published studies from which to draw technical specifications for team outcomes, and because most of these projects used a domain-dependent approaches to develop team models, the generalizability of those methods and technologies has not been realized.

This has been the major force behind the development of the Generalized Intelligent Framework for Tutoring (GIFT; Sottolare et al. 2011, 2012). GIFT is an open-source, modular architecture developed to reduce the cost and skill required to author, deliver, manage, and evaluate adaptive instruction. This framework is a powerful research tool and provides a starting point for a more advanced and flexible ITS processes. As part of its evaluation function, GIFT may be used as a testbed to understand the potential of ITSs as team tutoring tools (Sottolare and Holden 2013; Sottolare et al. 2011). In pursuit of this goal we began our research of team tutors by exploring the team and collaborative learning literature.

Group development might neatly be classified into three distinct areas of group interaction with different purposes: team training, teamwork, and collaborative learning (Cannon-Bowers and Bowers 2011). Van Berio (1997) compared and contrasted team training, team building, and cooperative learning. We have examined these differences and consolidated terms with similar terms in the literature:

- **taskwork team training** is a subset of team training which is focused on developing proficiency in task domains required for a specific duty of one's job (Salas 2015); taskwork team training is a domain-dependent learning activity; team training is often confused with the concept of teambuilding or teamwork (Van Berio 1997).
- **teamwork** is the "coordination, cooperation, and communication among individuals to achieve a shared goal" (Salas 2015, p.5); teamwork behaviors are largely domain-independent; teamwork includes the social skills needed to function as a team; teamwork activities may include teambuilding whose goal is to strengthen the coordination, cooperation, communication, coaching, conflict management, cohesion, and collective efficacy of the group (Salas 2015); teamwork is a necessary prerequisite to satisfactory taskwork performance (Van Berio 1997).
- **collaborative learning** (also referred to as cooperative learning) is "a situation in which two or more people learn or attempt to learn something together" (Dillenbourg 1999, p. 1); cooperative learning reinforces active participation (Van Berio 1997); collaborative learning generally focuses on a learning goal and is primarily domain-dependent and includes computer-supported collaborative learning (CSCL) activities.

While there are similarities between team taskwork, teamwork, and collaborative learning, there are also sufficient differences. This article focused primarily on teamwork and the identification of a set of behavior markers which indicate team states which are largely domain-independent. Behavioral markers indicating a high degree of collaboration for a group should not be confused with collaborative learning experiences. Although high collaboration within a group is usually an antecedent of high performance, collaborative learning experiences, which are focused on learning together, may be moderated by the group's ability to collaborate. Collaboration is an element of teamwork and collaborative learning is an instructional method to promote group learning. In this same way, a group of experts may each have a high degree of

proficiency in a particular task or collaborative learning domain, but may have their performance moderated by their ability to work together as a team. Teamwork is an antecedent of learning and successful performance (Van Berio 1997).

Teamwork Literature and Relevance to AIED and CSCL Research

Teamwork, team learning, and team performance seem to have sufficient coverage within the general training and education literature including the AIED literature on teams which seems to focus primarily on collaborative learning and collaborative problem solving. While the teamwork literature has received ample scholarly attention (Cannon-Bowers and Bowers 2011), little is known about the real ontology of core team behaviors, attitudes and cognition, and their influences on team outcomes (e.g., learning, performance, satisfaction, and viability). Most notably there have been few domain-independent approaches to the development of team models for computer-based tutoring. One such approach has its roots in neurophysiologic measures of collaboration (Stevens et al. 2009b, a; 2013). The focus of these team neurodynamics studies is to understand the changing rhythms and organizations of teams from the perspective of neurophysiology and specifically the concept of *neuronal synchrony*, in which a number, normalized between 0 and 1, quantifies the level of synchrony of a large population of neurons within a network or in this case individuals on a team. The theory suggests that higher synchrony measures between team members equate to higher team collaboration. Neurophysiologic measurement tools such as these may be a method to unobtrusive assessment of team states in team training experiences guided by ITSs.

Another such domain-independent approach is *cooperative learning*, where students work together to accomplish shared goals and are responsible to maximize not only their own learning, but the learning of all other group members (Johnson and Johnson 1986, 1999). In their ground breaking cooperative learning meta-analysis, Johnson et al. (2000) examined the impact of cooperative learning and compared several related learning strategies at the time. The primary independent variable in this study was the method of cooperative learning (a comparison of cooperation, competition, or individualistic learning) and the primary dependent variable was achievement as an outcome measure for performance.

The consistency of the results and the diversity of the cooperative learning methods provide strong validation for its effectiveness. However, the low number of studies conducted for several of the methods examined makes the reliability of their effect sizes very tentative. More and more studies have been conducted in the intervening years since the publication of the Johnson, Johnson & Stanne meta-analysis. Although this was not specifically addressed in the meta-analysis described herein, we recommend an update of the cooperative learning meta-analysis to strengthen the reliability of its results. Our goal in conducting the teamwork meta-analysis described herein was to expand the dimensions of teamwork (e.g., trust, cohesion, conflict management) to understand the broader influences for experiences beyond collaborative learning (e.g., team training and social interaction) and on outcomes beyond team performance (i.e., learning, satisfaction, and viability) where teamwork measures are represented by member attitudes, behaviors, and cognitions.

Others in the AIED and computer-supported collaborative learning (CSCL) community have built upon cooperative learning research to understand how individuals work together toward common goals and might be guided by ITS technologies (tools or methods) to enhance their learning, performance, or the quality of their overall instructional experience. Noteworthy articles by Erkens and Janssen (2008) and Dillenbourg and Hong (2008) discuss the assessment of dialogue acts during collaboration and the management of the collaborative environments respectively. A series of articles by McManus and Aiken (1993, 1995, 2016) highlight research in collaborative learning and collaborative problem solving. Soller (2001) adapted McManus and Aiken's (1995) Collaborative Skills Network (CSN) to form her Collaborative Learning Conversation Skill Taxonomy (CLCST) which included skills, subskills, attributes, and sentence openers to promote collaborative skills and support social interaction within the Intelligent Collaborative Learning System (ICLS). Since the behavioral markers identified in our meta-analysis are primarily verbal behaviors, it was logical to examine commonalities with CSN and ICLS prior to developing an implementation plan for team tutoring in GIFT. For example, negative markers identified in our meta-analysis could be cross-referenced with appropriate skills, subskills, or attributes in the CLCST and corresponding replies could be associated with sentence openers intended to mitigate negative behaviors and promote collaboration within the group.

As part of their research in "intelligent support for learning in groups," the AIED community examined the application of intelligent agents to: support peer tutoring (Walker et al. 2014); enhance collaboration (Adamson et al. 2014); trigger productive dialogue (Tegos et al. 2014); and assess groups during collaborative problem solving tasks (Rosen 2015). The research described in this article builds upon the AIED theme of intelligent support for learning in groups by identifying behavioral markers which may be used to assess various states of the team. Understanding the states and traits of teams and their progress toward individual and team goals is a necessary precursor for determining quality actions by the tutor and is also part of the motivation for understanding specific contributors to teamwork and team outcomes studied through the meta-analysis described in this article.

An understanding of collaborative behaviors might also be studied through team proxies where a premium is placed on communication and cooperation. Such is the case for interactions between human students and virtual peers managed by intelligent agents. Examples of this interaction is illustrated in studies of using dialogues, which assess interaction between one human student, one virtual student, and one virtual tutor (Lehman et al. 2013; Cai et al. 2014). Another example of a team proxy is the cooperative interaction described in Betty's Brain, where the student is responsible to teach the agent (Leelawong and Biswas 2008; Biswas et al. 2016). These agent-based studies offer insight to the tutor's perception-action coupling and highlight the need for the agent-based team tutor to be cognizant of the changing conditions of the environment and each of the individual learners on the team in order to sufficiently model the team, their interactions, and appropriate interactions by the tutor.

Finally, in our review of teams and teamwork in the AIED community, we examine the application of ITSs for a domain-specific team task. The Advanced Embedded Training System (AETS; Zachary et al. 1999) applied ITS tools and methods to the task of improving tactical training quality and reducing the

manpower needed for shipboard team training in the US Navy. AETS provided layers of performance assessment, cognitive diagnosis, and team-training support on top of an existing embedded mission simulation in the Navy's Aegis-class ships. Detailed cognitive models of the trainee's task performance were used to drive the assessment, diagnosis and instructional functions of AETS. The embedded training approach allowed tutoring functions to be integrated with training simulations implanted in the ship's equipment. This approach blurred the lines between training and work environments and while this approach was revolutionary and was expected to be leveraged across domains, this concept was not broadly applied and for whatever reason did not generalize to other training tasks. Regardless of its generalizability, this approach provided valuable lessons in how GIFT might be adapted to support team tutoring.

Drawing from the Sottolare et al. (2011) team tutoring model and others that tried to synthesize teamwork and team training in a qualitative way (e.g., Campion et al. 1993; Cannon-Bowers and Bowers 2011; Dyer 1984; Klein et al. 2009; Salas et al. 1992; Smith-Jentsch et al. 1998; Smith-Jentsch et al. 2008), we extracted key variables to scrutinize in a quantitative manner. Furthermore, since the relationship between these variables is complex, considering the different features and individual characteristics, intervention design, and environmental variables involved in team training, we recognize the team dynamism, and thus consider the ontologies for team outcomes in parallel to one another for clarity.

An Effective Team Architecture for ITSs

To develop a team ITS, additional work is required to identify a comprehensive design architecture, one delineating specific team model components, behavioral markers and associated measurement methods. This design architecture must be rooted in principles of intelligent tutoring, but also be based on the science of team performance and team training. Without a design architecture that contains concrete behavioral markers and assessment methods, it is difficult for a trainer or instructional designer to know how best to leverage team research to support collective training. In this article, our contribution is the discovery of significant antecedents to team performance and learning, and the identification of behavioral markers. Assessment methods will be evaluated in the future once these findings have been incorporated into GIFT.

In their summary of the literature, Dorsey et al. (2009) cited some of the main research challenges to be addressed to develop effective team ITSs were measuring team performance, improving team performance, and studying team formation and development. In light of this, we began a research initiative to develop an empirically-based ontology of the core attitudes, cognitions, and behaviors that influence team outcomes. These findings would then enable us to prioritize the most important factors influencing team outcomes that could then be instantiated in GIFT and validated in future GIFT-based tutors. This article describes the process and results of a quantitative synthesis of the existing science literature to inform the refinement of team models for performance and learning, and a process for applying these findings to ITS development. The results for satisfaction and viability outcomes will be published at a later date.

Methodology

Four meta-analytic structural equation modeling (MASEM) procedures were conducted to assess relationships of attitudes, cognitions and behaviors to team performance, learning, satisfaction, and viability. More importantly, these analyses contribute to literature on teamwork by providing an overarching model for each team outcome. This serves as a theoretical and practical understanding of teamwork within complex, diverse contexts by providing a more accurate nomological network, identifying gaps within the literature and practical insight for highlighting the simultaneous importance of team constructs, and opportunities for future research.

Literature Search

To identify primary studies for inclusion in the meta-analyses, a search was conducted using the American Psychological Association's PsychINFO (2003-July 2013), Defense Technical Information Center, and ProQuest for combinations and variations of the following keywords: performance/ competency/ trust/ cognition/ affect/ communication/ intelligent tutoring/ human-computer interaction/ virtual human/ mood/ emotion/ skill/ knowledge/ ability/ responsibilities/ roles/ distributed/ virtual/ after action review/ feedback/ leadership/ cohesion/ personality/ effectiveness; paired with either team/ unit/ group/squad/crew.

Furthermore, the following were used as secondary search terms: progress/ goals/ experience/ perceptions/ engagement/ boredom/ confusion/ frustration/ situational awareness/ training/ coordination/ collaboration/ motivation/ cohesion/ learning/ leadership/ training/ building monitoring/ goal setting/ instructional strategies/ debriefing/ decision making/ event-based training/ mental models (team, shared)/ processes/ shared cognition/ simulation based training/ development/ transactive memory systems/ back-up behavior/ planning/ coordination/ action/ transition.

Additionally, snowball, back-tracing approaches, and additional searches that included "team and learning"/ "teams and satisfaction"/ "teams and viability"/ "teams and performance" were used to supplement our searches.

In searching for primary studies, the search was bounded to include only those articles published/written during the 2003–2013 timeframe. This was done not only to make the model meaningful to current organizations (given the degree to which the nature of work has changed over the past 10 years), but also to complement and extend a number of meta-analyses which were published during the early 2000s (e.g., leadership by Burke et al. 2006; cohesion by Beal et al. 2003; team conflict by De Dreu and Weingart 2003). Our search yielded 5991 unique articles.

Inclusion Criteria

To be coded and included in analyses, articles needed to meet the following requirements. First, the study had to contain enough information to calculate a correlation between team variables to be included in the analysis. Second, the focus of the article had to be on teams whose members were interdependent. Third, due to a desire to focus on team performance within small teams, studies of teams that exceeded 9 people were not included. Finally, top management teams were excluded due to their unique nature

and dynamics. Shown in Fig. 1, the filtering method of inclusion/exclusion criteria resulted in a final meta-analytic database of approximately 300 primary studies, with 296 on team performance, 41 on team satisfaction, 18 on team viability, and 11 on team learning. This resulted in over 10,000 effect sizes prior to composites being created.

Coding Procedure

Studies that passed the inclusion/exclusion criteria were coded on several categories, including sample characteristics, reliability of measures, and effect sizes. A codebook was developed that provided detail on all of the components of the coding scheme to facilitate the quantitative coding process. Prior to beginning the actual coding each coder attended a team agreement meeting to ensure that the first 50 articles coded were consistent across coders in an effort to maintain inter-coder reliability. Each coder also received effect size and composite calculation training. Subsequently, pairs of coders were assigned articles whereby they came to consensus on which articles were deemed to be “codeable” (based on the boundary conditions specified earlier). Next, each article was coded by each individual in the pair and any discrepancies were resolved through a consensus meeting. To facilitate coding towards the end of dataset each pair of raters came to consensus on “codeability,” but then split those articles in half so that each individual coded one half of the identified articles.

Analysis Methods

For the quantitative analysis, we followed the Hunter and Schmidt (2004) guidelines for a random-effects meta-analysis of correlations. When multiple effect sizes were presented within a single sample, composites were created (Nunnally 1978), and if the information required to calculate a composite was not available, the mean of the effect sizes was used. In cases where a composite or average was calculated, the reported

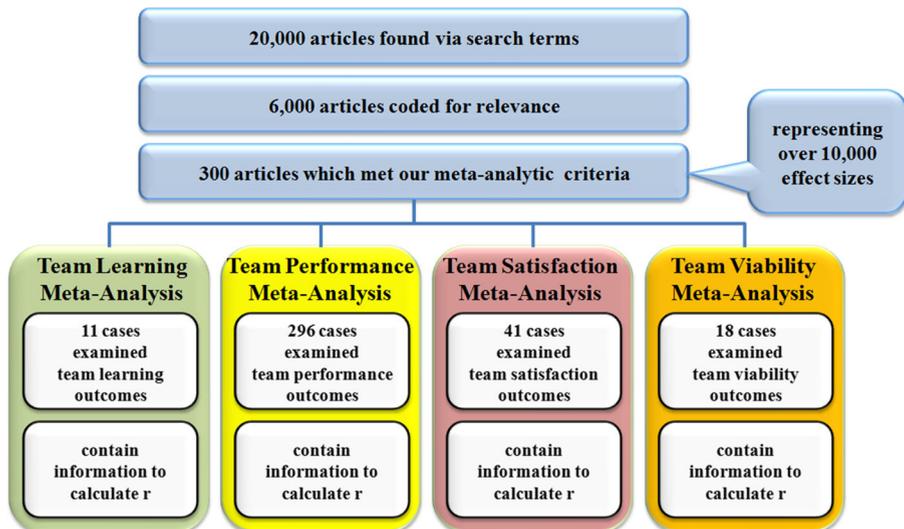


Fig. 1 Filtering method

reliability estimates were used in the Spearman-Brown formula (Li et al. 1996) in order to calculate the reliability of the composite or average. The calculation of the composite correlations and all analyses were performed using SAS Enterprise Guide 6.1 and SAS macros (Davis 2007) that executed original syntax as well as syntax modified from Arthur et al. (2001).

Our results included a sample-weighted mean point estimate of the study correlations (r) as well as the corresponding 95% confidence interval (which expresses the amount of error in r that is due to sampling error and is used for statistical significance testing). We also include the number of independent samples (k) and cumulative sample size (N) included in the calculation of the correlation (r_c) after correcting for unreliability in the predictor and criterion. Corrections for unreliability were performed using only the reliabilities reported in each article—no data was imputed.

In order to test the nomological network of team constructs, and therefore estimate their relative importance of these features for predicting team outcomes, a MASEM approach was applied. MASEM is a two-stage approach designed to test structural paths, providing a robust, theoretically driven quantitative review. We followed the recommendations from Viswesvaran and Ones (1995) and tested our model using LISREL 9 (Jöreskog and Sörbom 2004). First, we input the meta-analytic results into a correlation matrix.

Once we compiled all meta-analytic corrected coefficients, the harmonic mean was calculated to equate sample size varied across cells (see Table 2 for corrected coefficients). We draw from fit indices, such root-mean-square error of approximation (RMSEA), comparative fit index (CFI), non-normed fit index (NNFI or TLI), and reported chi-squared index (χ^2) with the caution that it is highly sample dependent to look for evidence whether the proposed model is adequate.

Due to widely recognized concerns about the restrictiveness of the χ^2 statistic, as well as its sensitivity to sample size (Jöreskog 1969; Quintana and Maxwell 1999), less sensitive and more reliable indices in assessing the reasonableness of the fit of the proposed models were also used in this study, including Tucker-Lewis Index (TLI), Comparative Fit Index (CFI), Root Mean Square Error of Approximation (RMSEA), and Standardized Root Mean Square Residual (SRMR; Ponterotto et al. 2003). The threshold values indicating good model fit are $> .95$ for CFI, $< .06$ ($N \geq 250$) for RMSEA, with $< .08$ as an upper limit (Hu and Bentler 1995), $< .1$ as an upper limit for SRMR, with $< .08$ suggesting excellent fit, and $> .90$ for TLI (Vandenberg and Lance 2000; Byrne 2001).

MASEM Results

Table 1 indicates the MASEM results for the primary team outcomes and all show adequate fit.

The desire to focus on the most relevant constructs for team tutoring resulted in limiting the results and discussion of findings to the *team performance* and *team learning* outcomes. While our findings for *team satisfaction* and *team viability* were within the 90% confidence interval (see Table 1, RMSEA), they exhibited lower confidence levels than for team performance and learning and were based on a much smaller number of studies. For these reasons, we decided not to include the full analysis

Table 1 Meta-Analytic Structural Equation Modeling (MASEM) Results

Ontology	χ^2	<i>df</i>	RMSEA (90% CI)	TLI (NNFI)	CFI	SRMR
Team performance ¹	4.306	3	.019 (.000; .056)	.994	.999	.008
Team learning ²	4.740	3	.040 (.000; .107)	.975	.996	.019
Team satisfaction ³	10.721	3	.081 (.033; .136)	.905	.981	.031
Team viability ⁴	2.080	1	.056 (.000; .166)	.988	.999	.007

Sample: $N_1 = 1173$; $N_2 = 350$; $N_3 = 396$; $N_4 = 346$; *RMSEA*, root mean square of approximation; *CI*, confidence interval; *TLI*, Tucker-Lewis index; *NNFI*, non-normed fit index; *CFI*, comparative fit index; *SRMR*, standardized root mean square residual

for either team satisfaction or team viability, but instead to focus on team performance and team learning. However, we have included the summary results for all four outcomes associated with the study in Table 1 for completeness.

Team Performance

Team performance is a primary focus of team research (e.g., Bell 2007; Cannon-Bowers and Bowers 2011). It has been defined as “the extent to which the productive output of a team meets or exceeds the performance standards of those who review and/or receive the output” (Hackman 1987, p. 323). According to our meta-analytic findings, team behaviors explain up to 42% of the variance in team performance. Considering the importance of distinguishing specific behaviors, we highlight action processes and organizational citizenship behaviors (OCBs) as the most important. These were followed by communication (13%), coordination (i.e., mutual support, 16%; reflexivity, 14%), leadership (11–17%), conflict management, transition processes, and conflict. A number of studies were found in this area of research, giving us a high degree of confidence regarding the moderate explanatory power of team behaviors to team performance (Table 2).

A wealth of research linking attitudes to team performance exists. The most commonly researched attitudes in relation to team performance were trust, collective efficacy, and cohesion. Results indicate that collective efficacy and psychological safety explain the most variance in team performance, 20% and 17% respectively. This is followed by trust and cohesion which also explain significant amounts of variance (9% and 15% respectively). Justice was also examined, but only had a k of 1 and was not significantly related to team performance (Table 3).

Team cognition also explained significant variance in team performance. Transactive memory systems and shared mental models accounted for 20% and 10% of the variance in team performance, respectively (see Table 4). Surprisingly, but with a caveat, situational awareness was the one construct to show the largest relationship with performance. However, this conclusion is based on small sample size that calls for further investigation in order to strengthen confidence in these findings.

These results are compared to those of the MASEM which allowed constructs to covary naturally. Findings from the MASEM indicated that the model below showed adequate fit ($CFI = .999$, $RMSEA = .019$, see Table 1). Figure 2 shows the integrative

Table 2 Behavioral contributions to team performance

Outcome	K^1	N^2	r_o	Rho^3	% Variance accounted
Overall	155	11,134	.25	.27	7%
Communication	64	4183	.31	.36*	13%
Coordination	49	3805	.26	.30*	9%
Mutual Support	11	881	.35	.40*	16%
Reflexivity	9	721	.33	.38*	14%
Monitoring	8	482	.08	.09	1%
Conflict	31	2062	-.14	-.17*	3%
Transition processes	4	235	.25	.30*	9%
Action Processes	4	345	.53	.65*	42%
Interpersonal processes	2	143	.19	.23*	5%
Leadership	46	3839	.29	.33*	11%
OCBs	4	313	.46	.55*	30%
Conflict management	10	902	.23	.27*	7%

¹ K is the number of effect studies analyzed for this outcome

² N is the number of individuals evaluated in this outcome

³ Rho is the corrected effect size, r_o is the uncorrected effect size; both weighted for sample size

*Statistically significant; confidence interval excludes zero

model that highlights the importance of collective efficacy ($\beta = .15$), cohesion ($\beta = .08$), communication ($\beta = .10$), and leadership ($\beta = .09$) behaviors. However, it also shows how most of the variance found in trust, coordination, and conflict influences performance through other mechanisms. Last but not least, the lack of inter-correlations present in the database does not allow for the inclusion of reflexivity, mutual support, OCBs, or any of the cognitive variables in the MASEM analysis.

Table 3 Attitudinal contributions to team performance: the effect of cooperation

Outcome	K^1	N^2	r_o	Rho^3	% Variance accounted
Overall	98	7587	.38	.44*	19%
Trust	23	1617	.26	.30*	9%
Collective efficacy	27	2657	.39	.45*	20%
Psychological safety	7	644	.33	.41*	17%
Cohesion	23	1540	.34	.39*	15%
Justice	1	170	.08	.09	1%

¹ K is the number of effect studies analyzed for this outcome

² N is the number of individuals evaluated in this outcome

³ Rho is the corrected effect size, r_o is the uncorrected effect size; both weighted for sample size

*Statistically significant; confidence interval excludes zero

Table 4 Cognitive contributions to team performance

Outcome	K ¹	N ²	r _o	Rho ³	% Variance accounted
Overall	39	2984	.32	.38*	14%
Team mental models	21	1478	.26	.32*	10%
Transactive memory systems	14	1206	.39	.45*	20%
Situation awareness	2	126	.56	.63*	40%

¹ K is the number of effect studies analyzed for this outcome

² N is the number of individuals evaluated in this outcome

³ Rho is the corrected effect size, r_o is the uncorrected effect size; both weighted for sample size

*Statistically significant; confidence interval excludes zero

Team Learning

For the purposes of this review, *team learning* has been defined as the acquisition of knowledge or skills through experience, practice, study, or by tutoring. It is important to understand what antecedents can foster learning, which has been highlighted as a core objective for any training intervention (Mesmer-Magnus and Viswesvaran 2010). According to our meta-analytic results, team behaviors account for 7–36% of the variance in team learning (Table 5). Specifically, conflict and conflict management appear as important antecedents, but the amount of studies available in this area limits the interpretation of such finding. Communication and reflexivity appear as main antecedents accounting for 25% and 15% of the variance, respectively. Additionally, these variables have a higher

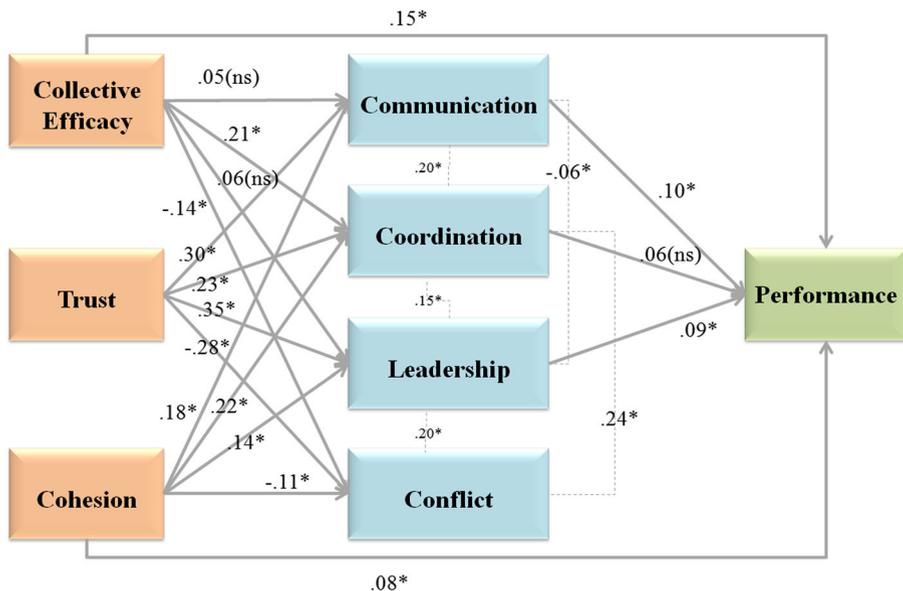


Fig. 2 Ontology for team performance based on meta-analytic structural equation modeling results

Table 5 Behavioral contributions to team learning

Outcome	K ¹	N ²	r _o	Rho ³	% Variance accounted
Overall	6	505	.51	.60*	36%
Communication	3	316	.44	.53*	28%
Coordination	3	343	.22	.26*	7%
Reflexivity	2	119	.33	.39*	15%
Conflict	2	209	-.37	-.44*	19%
Transition processes	1	78	.19	.23	5%
Action processes	1	78	.27	.33	11%
Conflict management	1	224	.59	.70	49%

¹ K is the number of effect studies analyzed for this outcome

² N is the number of individuals evaluated in this outcome

³ Rho is the corrected effect size, r_o is the uncorrected effect size; both weighted for sample size

*Statistically significant; confidence interval excludes zero

number of included studies thereby producing more confidence in results. Earlier work also suggests coaching/leadership can play a significant role in learning (e.g., Hackman and Wageman 2005; Edmondson et al. 2001).

Not surprisingly, attitudes reflective of cooperation account for 37% of the variance in team learning (Table 6). The role of psychological safety (74%), cohesion (44%), and trust (27%) becomes evident, but the issue of number of studies remains. The findings regarding team learning show a promising avenue that calls for future research to strengthen the confidence in the findings. There were no identified studies examining the relationship between team cognition and team learning.

A different picture emerges with the MASEM results where all antecedents are allowed to covary naturally. The model below showed adequate fit (CFI = .996, RMSEA = .040, see Table 1). Figure 2 shows the integrative model that highlights

Table 6 Attitudinal contributions to team learning: the effect of cooperation

Outcome	K ¹	N ²	r _o	Rho ³	% Variance accounted
Overall	5	523	.52	.61*	37%
Trust	4	307	.45	.52*	27%
Collective efficacy	1	78	.15	.18	3%
Psychological safety	1	178	.69	.86	74%
Cohesion	1	78	.56	.66	44%
Justice	N/A	N/A	N/A	N/A	N/A

¹ K is the number of effect studies analyzed for this outcome

² N is the number of individuals evaluated in this outcome

³ Rho is the corrected effect size, r_o is the uncorrected effect size; both weighted for sample size

*Statistically significant; confidence interval excludes zero

the importance of attitudes for cohesion ($\beta = .18$) and trust ($\beta = .11$), and conflict ($\beta = -.12$) and conflict management ($\beta = .27$) behaviors. However, it also shows how most of the variance that team behaviors (e.g., coordination, communication) account for with respect to learning occurs through other mechanisms. This highlights the importance of emergent states as antecedents when the goal is to improve learning. Moreover, the lack of inter-correlations among the data in our dataset limits the model in regards to the inclusion of reflexivity, psychological safety, leadership, or any of the cognition variables even though prior work has suggested these appear to be relevant.

Behavioral Markers

This section reviews the process for moving toward a set of behavioral markers for team performance and learning. Methodological choices when developing any measurement system are paramount for understanding human behavior (Meister 1985). This is especially true in intelligent tutoring systems (ITSs). In order to be designated an ‘intelligent’ system three conditions must be met (Burns and Capps 2013). First, the system must have a thorough understanding of domain knowledge in order to solve problems through the application of this knowledge. Next, the system must be able to accurately assess the knowledge of the learner within the system, and, finally, the system must be able to apply strategies to reduce the knowledge gap between the learner and the expert. This is the heart of the Learning Effect Model (LEM) on which GIFT is based (Sottolare 2012). Accurate measurement strategies are at the crux of all of these three steps. If measurement strategies are not thoroughly considered, an ITS will not be able to properly assess a learner’s current state, and, consequently, will not be able to engage in strategies to reduce the knowledge gap. This problem becomes compounded when the subject matter or domain is no longer declarative knowledge, but, rather, behavioral tutoring or training. In these instances, it is crucial to develop an accurate set of behavioral markers in order to provide the essential feedback needed in ITSs.

It is worth noting again that the meta-analysis described herein is intended to support a broad set of domains in which groups of people interact for a purpose. As noted earlier, collaborative learning is a prevalent theme in the AIED and CSCL literature (Rosé et al. 2001; Rosé and VanLehn 2005; Kumar et al. 2007; Erkens and Janssen 2008; Chaudhuri et al. 2009; Kumar et al. 2010; Ai et al. 2010, Kumar et al. 2010; Kumar and Rosé 2011; Adamson et al. 2013; Dyke et al. 2013; Adamson et al. 2014), but groups also interact for other purposes (e.g., to enhance task performance or to develop social skills as in teamwork). The behavioral markers identified herein provide a mechanism to identify teamwork attributes beyond collaboration.

Presently, our goal is to develop a set of markers that can be used for an intelligent team tutoring system. ITSs utilize artificial intelligence to select and deliver directed feedback during training situations. Traditionally, the focus of ITSs has been on the development of an individual’s cognitive skills (e.g., problem solving and decision making). Yet, this dismisses the potential to leverage these types of systems to create opportunities for social learning in team environments (Singley et al. 1999). Thus, the current objectives were focused on developing behavior markers specifically in the team context. In order to do this, we first utilized a refined team model. Specifically, we reviewed the current state of the team literature and, using the initial GIFT team models

(Sottillare et al. 2011; 2012) as basis, we created a contemporary team model architecture. This design architecture was then used to identify crucial team states that would be the foundation of selected behavioral markers.

Next, we briefly describe the most common methodology used to assess team function – use of self report Likert scales – its weaknesses and why the use of behavioral markers is more functional when trying to develop an ITS architecture for teams. We will then move to describing the process used to develop an initial set of behavioral markers for a limited number of the team constructs identified as being important in our meta-analytic research.

Why Use Behavioral Markers?

Measurement techniques are a crucial component when considering the validity and generalizability of scientific research. When designing a measurement system, one must make several decisions considering the pros and cons of measurement source (e.g., supervisor, self-report, trained rater), when to measure the construct of interest (e.g., beginning, during, end of performance episode), or what scaling technique should be used (e.g., Likert, behavioral markers; paired-comparisons). Traditionally, one of the most common techniques used has been self-reported Likert-like measures taken at the end of the performance episode.

However, there are some underlying criticisms of using the self-report measures as a main approach. First, self-reported measures are more subject to social desirability bias than other measurement sources (Budescu and Bruderman 1995). In other words, individuals have a tendency to exaggerate estimates concerning the positive characteristics when they are the referent of the measure. Second, measurements taken at the end of a performance episode are more susceptible to judgment errors like availability heuristics. The availability heuristic suggests that the qualities of an event are influenced by the ease with which one can recollect the event (Tversky and Kahneman 1974). As such, more emotionally-laden events will be recalled a lot easier and, subsequently, determined to occur more frequently – even if they did not occur often. Lastly, the scale points used in Likert-like measures (e.g., 1 = *Strongly Disagree*; 7 = *Strongly Agree*) are subject to interpretation by the individual filling out the measure. That is, there might be different standards one uses to judge each scale-point and these judgments may be influenced by pre-existing stereotypes (Biernat and Manis 1994). However, there has been movement to reconcile some of these issues using different measurement techniques.

Recently, researchers have started to move towards a more objective and less obtrusive approach when measuring psychological constructs (Wiese et al. 2015). This approach uses an objective source (e.g., trained rater, intelligent system) that makes judgments on a set of behavioral markers during a performance episode. This approach reduces that amount of error compared to self-reported Likert measures taken at the end of the performance cycle. More specifically, using more objective sources reduces the degree of social-desirability bias, availability heuristics may be limited when ratings are made during performance episodes, and using behavioral markers removes the possibility of changing standards between participants. As such, using this measurement technique will result in less biased and more accurate judgments than using the type of self-reported Likert-type measures which are often collected at the end of a study.

However, we acknowledge that objective measures also have their limitations with reported issues such as criterion deficiency or contamination (Borman and Smith 2012; Schmidt and Hunter 1977) or potential biases (Bound 1989).

Deriving Behavioral Markers

In order to develop the behavioral markers, we had to first identify critical team constructs that were indicators of effective team performance. Using the results of the meta-analytic investigation reported earlier, we refined the team states included in the initial GIFT Framework (Sottolare et al. 2011). We then took a subset of these constructs (i.e., psychological safety, trust, communication) that were shown to be predictors of team outcomes (e.g., learning, performance, viability, and satisfaction) and searched the literature for measures that have been used to assess these constructs. Next, we compiled the items from these measures into separate excel documents. These excel documents were then given to several subject-matter experts (SMEs) with the goal of (1) developing behavioral markers that represented each individual item, (2) identifying and removing items that measured the same behavior, and (3) developing sub-dimensions of the construct as a whole if necessary.

Once the SMEs completed their ratings, the group met and came to a consensus regarding whether the generated markers accurately represented the item. During this meeting, SMEs determined whether the construct and its sub-dimensions were equally represented by the markers. If the group believed that the current markers were deficient in representing any aspect of the construct, new markers were then generated and agreed upon. This was the final step in generating the final list of behavioral markers by construct. A shortened version of these steps is displayed in Table 7. Following this process, we identified markers for antecedents that accounted for significant variance in one or more of our outcomes. Next we describe potential markers for trust, collective efficacy, conflict, conflict management, and communication. It should be noted that while the behaviors listed are domain-independent, the markers represent strategies to recognize team behaviors in any domain. It is apparent that in a specific domain under training, there would need to be domain-specific measures to understand their context or the frequency of occurrence.

Table 7 Marker identification process

Step	Action
1.	Identify Crucial Team Construct related to Team Outcomes
2.	Search the literature for measures of these constructs
3.	Compile items from for each team construct
4.	Independently Review Items
4a.	- Create markers for each item
4b.	- Identify duplicates
4c.	- Come up with sub-dimensions
5.	Meet as a group to come to an agreement on markers and sub-dimensions
6.	Identify areas of sub-dimensions that may have a deficient markers
7.	Generate new makers to fill in deficiency

Trust

While there is no universally agreed upon definition of trust, two elements that are common to nearly all definitions reflect the idea of positive expectations and the willingness to become vulnerable. Perhaps one of the most cited definitions argues that trust is, “the willingness of another party to be vulnerable to the actions of another party based on the expectation that the other will perform a particular action important to the trustor, irrespective of the ability to monitor or control that other party” (Mayer et al. 1995, p. 712).

A search of the AIED literature for the term “trust” produced no results, but a search of “expectations” did produce a relevant article that examined the tailoring of progress feedback and emotional support related to the personality trait “agreeableness” (Dennis et al. 2016). They indicate that team members with higher levels of agreeableness exhibit higher levels of trust. They therefore advocate a strategy of providing advice and reassurance to learners with low performance and high agreeableness. In contrast, they advocate providing with only advice for moderate and high performers with high agreeableness. However, this model only covers a propensity for trust and does not include additional antecedents of team trust put forth by Costa (2003). These include the preference of individuals for working on teams, adequate job skills of team members, tenure within the team, cohesion of the team, functional dependence of the task or process being executed by the team, the perceived trustworthiness of other members, and cooperative and monitoring behaviors.

In this vein, below we present several trust markers. More specifically, we present a mixture of markers some of which describe the trustee’s behavior (the target) and the trustor (person trusting). The former is typically indicative of trustworthiness and the later trust. Items here generally fall into markers which reflect assessments of competence, benevolence, or integrity. Some of which are surface level and some are more deep level (only appearing after a time). The markers in Tables 8 and 9 below were created based on items that appear within the following published papers on trust (Chang et al. 2012; Cook and Wall 1980; De Jong and Elfring 2010; Dirks 2000).

Table 8 Trust markers

-
1. Frequency which team members remind each other to follow through on their tasks (reverse coded)
 2. Frequency which team members accept back up behaviors from their fellow teammates
 3. Occurrences of team members actively seeking out the opinions of other team members regarding work tasks
 4. Amount of task information withheld from fellow team members (reverse coded)
 5. Occurrences of team members delegating task responsibilities, assigning additional tasks, or asking the advice of their fellow team members
 6. Occurrences of team members “using honest language” (e.g., I’m just being honest with you) in discussions (judgment; reverse coded)
 7. Occurrences of team members double checking their fellow team members work or calling their competence into questions in a public forum (i.e., with their fellow team members; reverse coded)
 8. Occurrences of team members pointing out their fears, providing updates on the quality of information, questioning competence, or other potentially negative information amongst team members. (e.g., questioning team members; reverse coded)
-

Table 9 Trustworthiness markers

-
1. Providing negative feedback in an inappropriate way (e.g., involving interpersonal insults) or at an inappropriate times (e.g., in front of other team members; judgment; reverse coded)
 2. Frequency of team members completing task responsibilities in a timely/assigned manner
 3. Occurrences of team members engaging in inappropriate interpersonal behaviors (e.g., cutting people off, tell them to ‘shut up,’ name calling) during any part of the task (reverse coded)
 4. Occurrences of team members providing either good verbal or physical backup behaviors
 5. Occurrences of team members purposefully providing bad information (reverse coded)
-

Collective Efficacy

Collective efficacy has been defined as, “a shared belief in a collective’s capabilities to organize and execute the course of action” (Bandura 1997, p. 477). In essence, a team experiences a sense of collective efficacy when they perceive that the team members within the team have the strong abilities to fulfill their roles on the team as it relates to the task in question. Interpersonal factors can moderate the relationship between perceived abilities and task perception such that when interpersonal factors are diminished (e.g., personal conflict is high), teams may feel they have less collective efficacy not due to task-related capabilities but interpersonal factors which impact task execution. Scales on collective efficacy tend to be task-specific and incorporate explicitly task-driven factors or may also include some of the interpersonal factors (given that both task work and teamwork are needed for a team to successfully perform). Therefore, the markers for collective efficacy reflect this duality: ability and task domain context. In this sense, collective efficacy is a form of trust in the abilities of the team.

A search of the AIED literature for the terms “collective efficacy”, “collective worth”, “collective value” and related terms produced no results. The collective efficacy markers were created based on items that appear within the following published papers on collective efficacy (Bray 2004; Chen et al. 2005; Cheng and Yang 2011; Cronin et al. 2011; Edmonds et al. 2009; Guzzo et al. 1993; Hsu et al. 2007; Jehn et al. 1997; Jones 1986; Lent et al. 2006; Luhtanen and Crocker 1992; Mathieu et al. 2009; Riggs and Knight 1994; Sargent and Sue-Chan 2001; Shivers-Blackwell 2004; Tasa et al. 2007; Woehr et al. 2013; Wong et al. 2009). Within this set of markers we have the most confidence in the ability-focused markers. For brevity only a sampling of the 19 ability-focused markers and 13 contextualized markers discovered are displayed in Tables 10 and 11. A full list is available from the authors upon request.

Table 10 Ability-focused collective efficacy markers

-
1. Team members actively request backup when needed
 2. Team members acknowledge input from teammates during taskwork, and incorporate input into team task
 3. Team members acknowledge fellow teammates’ inputs and suggestions
 4. Team communication portrayed positive messages, such as reinforcing team ability to complete task, avoid negative or anxiety of failure, etc.
 5. Team members request assistance from/direct questions to a team member rather than an external member or external instructor
-

Table 11 Contextualized collective efficacy markers

-
1. Express confidence and willingness to resolve disagreements
 2. Team members cooperate with one another and make effort to avoid conflict
 3. Team members expressed confidence that they could efficiently resolve conflicts regarding opinions, preferences, differences, etc. about how to complete the task and begin working toward this goal
 4. Team members expressed confidence that they could efficiently resolve conflicts regarding which tasks to prioritize and complete the task and begin working toward this goal
 5. Manage interpersonal and/or task conflict effectively while prioritizing team tasks (discussion of task, execution of team task-related actions)
-

Cohesion

While cohesion has been defined in many ways, it has generally been argued to reflect an attraction or bond within the team. More specifically, cohesion has been defined as “the bonding together of members of a unit in such a way as to sustain their will and commitment to each other, their unit, and the mission” (Johns et al. 1984, p. 4). Carless and De Paola (2000) conceptualize cohesion as having three dimensions: social, task, and group pride, as such we have examined the existing scales and papers which have reported to have cohesion measures in them and delineated a first round of markers that tap each of the three dimensions. In this vein, task cohesion is defined as the group’s shared commitment and ability to execute the group task or goal, or the group’s capacity for teamwork (Siebold 1999; Craig and Kelly 1999). Social cohesion reflects the group’s bond that promotes development and maintenance of social relationships (Carless and De Paola 2000). Finally, group pride can be described as a shared sense of unity and honor derived from membership in the group and the group’s accomplishments.

A search of the AIED literature for the term “cohesion” and “commitment” produced a relevant article that examined the goal of verbal and non-verbal communication. Communications were classified as either “aimed at solving a problem or alternatively aimed at creating social cohesion and team spirit.” (Rosenberg and Sillince 2000, p. 299). Communications that were aimed at problem solving (task oriented) tended to suppress social meanings and render them invisible and communications that were aimed at social cohesion tended to suppress task meanings and render them invisible. This research contributes directly to both task cohesion (Table 12) and social cohesion (Table 13).

The markers in Tables 12, 13 and 14 were created based on items that appear within the following published papers on team cohesion (Carless and De Paola 2000; Carron et al. 1985; Chang and Bordia 2001; Henry et al. 1999; Hobman and Bordia 2006; Hoegl and Gemunden 2001; Jehn 1995; McCroskey and McCain 1974; Miller 1964; O’Reilly et al. 1989; Podsakoff et al. 1997; Rosenfeld and Gilbert 1989; Rosenberg and Sillince 2000; Sargent and Sue-Chan 2001; Shin and Choi 2010; Shivers-Blackwell 2004; Solansky 2011; Watson et al. 1991; Wong 2004; Zaccaro 1991). For brevity, only a sampling of the 29 task cohesion markers and 16 social cohesion markers are shown in Tables 12 and 13 respectively, but all of the 5 group pride markers discovered are displayed in Table 14. A full list is available from the authors upon request.

Table 12 Task cohesion markers

-
1. Members are actively working together and pitching in to reach team goals
 2. Contributing to discussions about new course of action/problem solving
 3. Owning up to mistakes when they occur
 4. Valuing member contributions; seeking input from all members
 5. Occurrences of phrases like “we make a great team”, acknowledging each other’s strengths and how they complement weaknesses in the group at large
 6. Occurrences of phrases like “great job everyone”, “go team”, “you’re the best”, “good work”; positive affirmations toward the team’s work
-

Communication

Communication has been defined as, “the process by which information is clearly and accurately exchanged between two or more team members in the prescribed manner and with proper terminology; the ability to clarify or acknowledge the receipt of information” (p. 345, Cannon-Bowers et al. 1995). In examining published measurement instruments that pertain to communication we found that when thinking about communication markers one must be cognizant that team communication is a complex endeavor and consideration must be given to the application of these markers and potential measurement methods. The characteristics of the communication as well as the content of team communication (see Table 15) are important to determining the effectiveness of team communications.

A search of the AIED literature for the terms “team communication” and “group communication” produced no results, but a search of “communication” produced several relevant articles that included titles related to learning in groups (Kumar and Kim 2014; Yoo and Kim 2014; Walker et al. 2014; Tegos et al. 2014; Adamson et al. 2014), social relationships and responsibilities in AIED systems (Walker and Ogan 2016), and peer collaboration in distributed learning environments (Greer et al. 1998; Muehlenbrock et al. 1998). While important mediators of team communications, the focus of these articles in most cases was to facilitate discussion among a group of peers rather than determine whether the group’s communication indicated progress toward a team goal.

The markers in Tables 16, 17, 18 and 19 were created based on items that appear within the following published papers on communication which contain scales (Bunderson and Sutcliffe 2002; Cronin et al. 2011; De Dreu 2007; Espinosa et al. 2012; Faraj and Sproull 2000; Fletcher and Major 2006; Gajendran and Joshi 2012; Greer et al. 2012; Gupta et al. 2009; Haas 2006; Hirst 2009; Huang and Cummings

Table 13 Social cohesion markers

-
1. Team members rarely socialize outside the work context (reverse coded).
 2. Number of social events organized by the team outside of work.
 3. Number of social events voluntarily attended.
 4. Team members rarely share personal stories (reverse coded).
 5. Introducing and/or mentioning each others’ names [first or last, or rank as appropriate].
-

Table 14 Group pride markers

1. Messages of positive affirmation about being part of the team, usually made outside the immediate group.
2. Desire to stay in the group; come back and train with same people (team viability).
3. Coming up with a team name, call signs, patches, t-shirts, logos.
4. Sticking up for the team, defending the team and its members.
5. Highlighting the team's accomplishments

2011; Jong et al. 2005; Lee et al. 2010; Schippers et al. 2007; Tung and Chang 2011; Walther and Bunz 2005).

Tables 16, 17, 18 and 19 provide a set of potential markers for identifying the effectiveness of team communication. The tables highlight four sub-dimensions of communication: 10 general information sharing behaviors (Table 16), 17 contextualized information sharing behaviors (Table 17), 1 workflow knowledge sharing behavior (Table 18), and 1 characteristic of inter-team (team-to-team) communications (Table 19). While these markers are expressed as domain-independent behaviors, the related measures of these behaviors would be specific to their domains (e.g., task relevant information sharing would require identification of unique information for each task domain). For brevity, only a subset of the general information and contextualized sharing behaviors are shown.

Conflict & Conflict Management

Team conflict has been defined as, “the process resulting from the tension between team members because of real or perceived differences” (De Dreu and Weingart 2003, p. 741). Closely related is the notion of conflict management which refers to the process

Table 15 Team communication characteristics and content

Characteristic	Content
Task/Non-Task	Transition phase
Concise	Mission analysis
Timely	Goal specification
Clear	Strategy formulation
Closed-Loop	Action phase
Push vs. Pull	Monitoring goal progress
Unique vs. Shared	Systems monitoring
	Team monitoring/backup
	Coordination
	Interpersonal
	Conflict management
	Motivation/confidence building
	Affect management

Table 16 Team communication markers for general information sharing

-
1. Occurrences of task relevant information being shared
 2. Team member communication is concise
 3. Team member communication is timely
 4. Team member communication is clear
 - a. Frequency of requests for clarification (reverse coded)
 5. Frequency with which team members engage in closed loop communication
 - a. Team members verbally acknowledging when a team member shares information or makes a suggestion
 - b. Frequency with which team members confirm receipt of information presented to them (e.g., acknowledgment) by other team members (e.g., Okay, Got it, Roger that, etc)
 - c. Frequency of closed loop sequences (task specific)
-

through which team members engage in strategies to effectively manage conflict as it arises.

A search of the AIED literature for the terms “conflict” and “conflict management” produced key results (Tedesco 2003; Hall et al. 2015; Israel and Aiken 2007). These approaches both classify/categorize and mediate conflict by guiding the group toward cooperative behaviors through constructive discussion in specific domains. The goal for the meta-analysis described herein is to provide generalized behavioral markers to classify/categorize conflict. Once the behavior is classified, generalized strategies (which are yet to be developed) would be applied to mediate the conflict. While these approaches are domain-dependent, the processes used to identify mediation strategies are relevant to next steps in identifying generalized (domain-independent) strategies.

In examining published scales reported to measure conflict, we found a mixture of conflict and conflict management markers. Therefore, we have included both below. The markers in Tables 20, 21, 22 and 23 were created based on items that appear within the following published papers on conflict (Barker et al. 1988; Espinosa et al. 2012; Gupta et al. 2010; Huang 2010; Jehn 1995; Jehn and Mannix 2001; Kostopoulos and Bozionelos 2011; Pearson et al. 2002; Rispens 2012; Rutkowski et al. 2007). For brevity, only a subset of the 18 conflict management markers discovered are shown (Table 23).

Table 17 Team communication markers for contextualized information sharing

-
1. Occurrences of team members clearly stating their expectations regarding the main mission tasks (transition phase)
 2. Occurrences of the team sharing information regarding environmental conditions (transition phase)
 3. Occurrences of team members discussing roles and responsibilities (establishment of, sharing information pertaining to who has a role and its capabilities/responsibilities; transition phase)
 4. Occurrences of team members updating one another on progress towards goals (e.g. what remains to be accomplished, current progress; action phase)
 5. Occurrences of team members providing verbal feedback to one another (action phase)
 6. Occurrences of information being shared regarding the sequencing of team member action (action phase)
-

Table 18 Team communication markers for workflow knowledge sharing

-
- Frequency of occurrences of task knowledge being requested from each team member (this would be more of a social network measure)
-

Application of Findings to GIFT

The behavioral markers identified in this article shaped foundational measures needed to support team tutoring not only in GIFT, but in other ITS architectures. The ontologies shown in Figs. 2 and 3 were derived from the MASEM process and shape a set of initial models of team performance and learning respectively. Next steps in applying these findings to GIFT will be to validate these models through a program of rigorous experimentation across several task domains and populations.

Implementation within GIFT will require modifications to nearly all aspects of the framework and its underlying theoretical foundation, the Learning Effect Model (LEM; Sottolare 2012; Fletcher and Sottolare 2013, Sottolare et al. 2013) which has been updated as shown in Fig. 4. The LEM illustrates the relationship between the common elements of ITSs: learner model, pedagogical model, and domain model. The GIFT ontology uses the term “module” for major elements of the architecture because each module contains both a model and active software.

In the LEM the learner model (green boxes) is composed of real time learner data (derived from sensors and learner input), learner states (derived from learner data and classification methods), and long term learner attributes derived from the results of previous inputs and experiences. The long-term model represents a record of the learner’s demographic data and historical data (any relevant data including achievements), and a short-term model representing any acquired learner data (e.g., data from long-term learner model, real-time sensors or learner input or actions) and derived learner states (e.g., domain performance, cognitive, affective or physical). The pedagogy (light blue boxes) is represented by instructional strategies and agent policies. The pedagogy is domain-independent as instructional strategies include the ITS’ plan for action (e.g., prompts, hints, and questions) and agent policies are based on best instructional practices from the literature (e.g., mastery learning, error-sensitive feedback, and fading worked examples). The domain (orange boxes) includes instructional tactics selection and environmental conditions.

Instructional tactics (actions by the tutor) are influenced by strategies selected by the pedagogical module in GIFT. If a “prompt learner for more information” strategy is selected by the pedagogical module, then the domain module selects a tactic that is specific to that domain and the environmental conditions. GIFT uses the term “environmental conditions” to broadly represent the domain content. In a mathematics tutor, the environmental conditions would be specified as the problem type and difficulty level (e.g., moderately difficult quadratic equation), but in an immersive virtual

Table 19 Inter-team communication markers

-
- Similar items as within information sharing, but add the caveat that the information exchange is about exchanging information outside the team
-

Table 20 General conflict markers

-
1. Frequency of disagreements over performance episode(s)
 2. Frequency of disagreements over performance episode(s)
 3. Have substantial agreement on goals, strategies and processes (reverse coded)
 4. Nonverbal signals (furrowed brow, loud volume, red face) (reverse coded)
 5. Team member verbally communicates frustration with team (e.g., phrases, loud sighs, or an unwillingness to continue working with someone or on a task) (reverse coded)
 - a. Team members raise their voice at each other (reverse coded)
-

simulation, the environmental conditions would represent discrete states (e.g., state of entities at a specified point in a training scenario). In this way, GIFT can be applied broadly to almost any domain. However, at this time, we cannot support collective (team) instruction. This is the objective for developing a team process for tutoring.

Notionally, the LEM for individual learners (Fig. 4) will be modified as shown in Fig. 5 to support effective tutoring for teams. Learner models in GIFT will be expanded to support multiple individual learner models for each team member and a multi-dimensional team model will need to be added. Note that the team data includes the behavioral markers identified for team performance and learning antecedents. The GIFT inter-module message set, based on ActiveMQ, will be expanded to support sharing of team state information. The domain module in GIFT will also be modified to accommodate team concepts (learning objectives) and associated measures, and the GIFT authoring tool (GAT) will be updated to allow GIFT ITS developers to define required team knowledge and skills, associated team learning and performance objectives, team instructional experiences (e.g. sets of problems to solve collaboratively or immersive simulation scenarios), and associated methods to acquire and interpret team measures.

In addition, the presence of multiple team members presents new feedback choices. As noted by Bonner et al. (2016) and Walton et al. (2014), all team members might receive feedback (“Team, we need to...”), all team members might receive individuals’ feedback with names attached to re-engage learners (“Alice, remember the goal is...”), individuals might receive private tailored feedback, or a subset of the team might receive feedback. This choice in feedback design dramatically affects both the pedagogy of the learning environment and the volume and complexity of feedback that members receive.

Team members’ cognitive load may be increased if receiving feedback requires the dual task of processing both the content of the feedback and the impact of receiving that feedback. However, even in the cognitively simpler mode, in which all feedback messages are addressed to “Team” and received by everyone, no matter which member’s behavior triggered them, the sheer number of messages received increases exponentially with the number of team members. Thus, the filtering of communications shown at left in Fig. 5 and managed by the instructional tactics selection process is critical in intelligently managing the flow and frequency of feedback to the team.

Table 21 Interpersonal conflict markers

-
1. Frequency of disagreements about opinions
 2. Ratio of task to relationship conflicts, taking into account frequency of disagreements
-

Table 22 Task conflict markers

-
1. Frequency of disagreements about which tasks have highest priority during performance episode(s)
 2. Frequency of disagreements about who should do what over performance episode(s)
 3. Frequency of disagreements about who should receive what resources over performance episode(s)
 4. Frequency of disagreements about who should do what over performance episode(s)
-

Choosing which feedback messages to pass through to the learner should be based on their priority, on how many communications the learner has recently received, how much feedback other members have received (to promote equity), and even on the individual learner's cognitive processing capabilities and current cognitive task load.

It is also worth noting that the team architecture in Fig. 5 not only doubles its observe-assess-respond loop from individual to team, but also doubles the internal storage of almost all its components, to monitor both team skills and task skills. Thus, the Long Term Team Model will contain data about a particular team's ability to communicate as a team and its ability to perform specific team tasks, such as conducting a patrol. The Long Term Learner Model for the individual will contain data about a specific member's skills within a team task as well as data about his/her ability to cooperate with others on a team, for example. The instructional strategy selection within the pedagogical module will now take team dynamics into consideration along with task performance (at, above, or below expectations) as strategies are selected. In effect, when we train teams, we aspire to teach members not only how to perform the task but how to effectively and efficiently work as a team. Thus, we are teaching much more content to our learners than previously in a similar timeframe, and must develop the architecture to accommodate this additional teaching.

Applying the MASEM Analysis to the GIFT Architecture

In moving from individual to team tutoring, GIFT will need to be adapted to support not only the assessment of concepts (e.g., learning objectives), but also the assessment

Table 23 Conflict management markers

-
1. Team member communicates the benefits of his/her/their opinion, strategy or approach
 2. Team member or leader encourages a 'we are in it together' attitude
 3. Team member(s) or leader seeks a solution that will be good for all of us
 4. Team members adhere to current plans/opinions/etc. (e.g., stick to protocol) until teammates provide reasonable rationale otherwise for changing
 5. Team members are discouraged from showing anger or losing their cool (e.g., punishment)
 6. Team members are open and receptive to differing opinions, strategies or approaches
 7. Team members are respectful of one another (no name calling, no talking down to team members)
 8. Team members disclose instances of conflict/tension to individuals outside of the conflict (e.g., leaders, supervisors, other teammates)
 9. Team members identify conflict
 10. Team members initiate addressing conflict when it arises
 11. Team members refrain from showing anger that may arise, use strategies to keep one's self calm
-

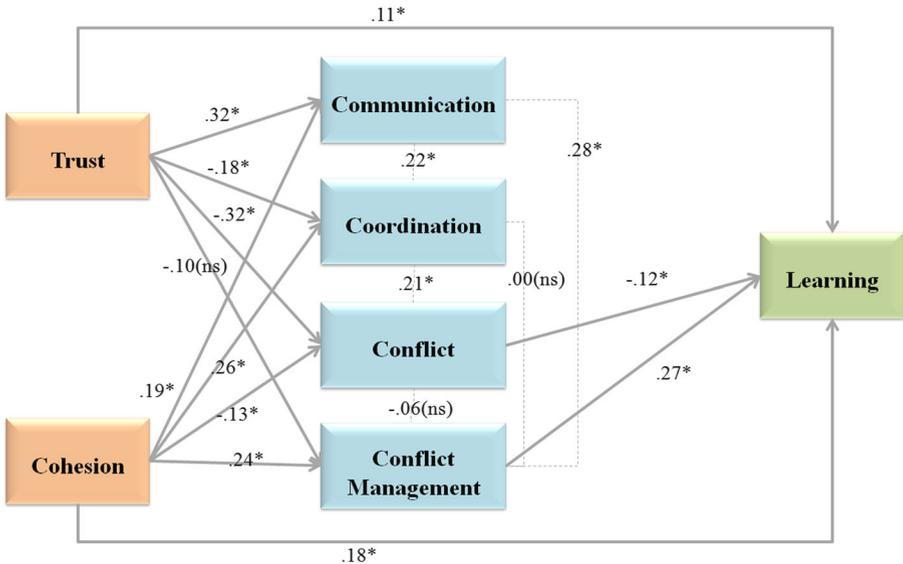


Fig. 3 Ontology for team learning based on meta-analytic structural equation modeling

of teamwork concepts. Examples of teamwork objectives based on our meta-analysis might include high trust, trustworthiness, collective efficacy, and cohesion, along with timely and relevant communications, low conflict, and rapid resolution of conflict when it does occur. We consider the probability to generalize teamwork objectives, measures, and remediation by the tutor across domains as high based on the markers identified in both our meta-analysis and those found by Johnson et al. (2000) and implemented as sentence openers by Soller (2001).

By way of example, the following is offered as a detailed trace of changes needed in GIFT enumerations to support team tutoring. Trust accounts for 27% of the variance in team learning and has behavioral markers which include “the amount of task information withheld from fellow team members” indicating distrust. In a given team tutoring scenario, the amount of task information available and not disclosed can be compared to the total information available. In order for a GIFT-based tutor to assess trust based on this single measure, a public class would need to be created in JavaScript (shown

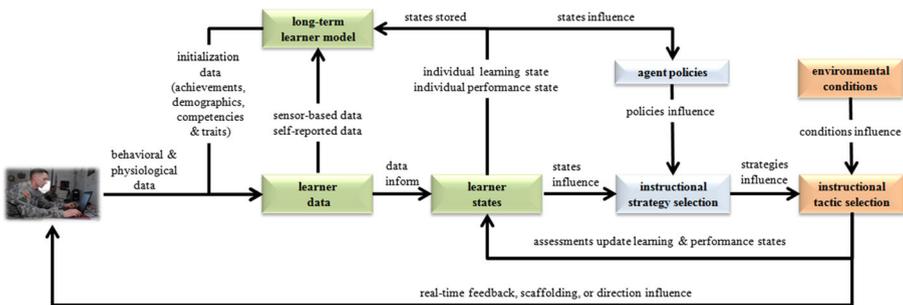


Fig. 4 Updated Learning Effect Model (LEM) for individual learners: GIFT learner module (green boxes); GIFT pedagogical module (light blue boxes); GIFT domain module (orange boxes)

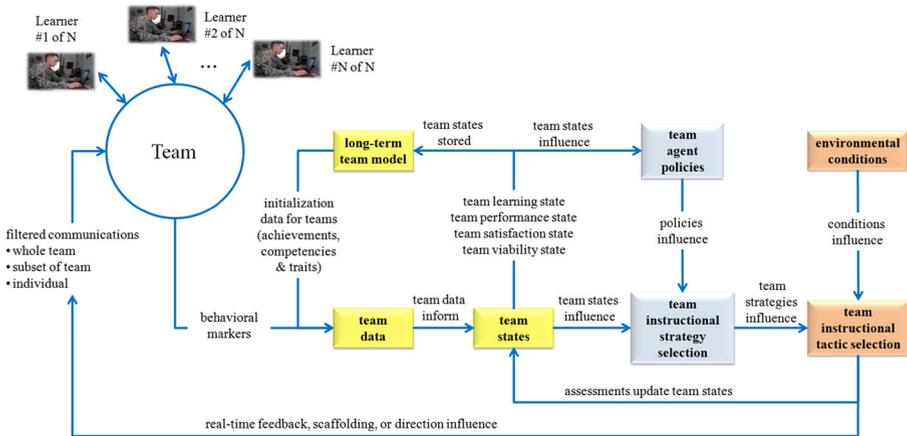


Fig. 5 Learning Effect Model (LEM) for teams: GIFT team module (yellow boxes); GIFT team pedagogical module (blue boxes); GIFT team domain module (orange boxes)

below) where low, medium, and high trust enumerations would be equated to the number of occurrences of one or more of the markers identified in Tables 8 and 9.

```

public class TrustLevelEnum extends AbstractEnum {

    private static final List<TrustLevelEnum> enumList = new
    ArrayList<TrustLevelEnum>(2);
    private static int index = 0;

    public static final TrustLevelEnum UNKNOWN = new TrustLevelEnum("Unknown",
    "Unknown");
    public static final TrustLevelEnum LOW = new TrustLevelEnum("Low", "Low");
    public static final TrustLevelEnum MEDIUM = new TrustLevelEnum("Medium",
    "Medium");
    public static final TrustLevelEnum HIGH = new TrustLevelEnum("High",
    "High");

    private static final TrustLevelEnum DEFAULT_VALUE = UNKNOWN;
    
```

This same type of class definition is required for all behavioral markers along with a defined method of data acquisition and a classification strategy (e.g., rules, decision trees).

The changes required for GIFT to support team tutoring and specifically teamwork measures will include changes to the engine for managing adaptive pedagogy (eMAP; Fig. 6). EMAP is the default strategy engine in GIFT and is based on another extensive review of the training literature (Goldberg et al. 2012) to determine best instructional practices for individual instruction. Figure 6 shows the authoring tool interface for configuring pedagogical relationships between individual learner attributes (e.g., motivational level), the quadrant of instruction (rules, examples, recall or practice according to Merrill 2015), and metadata tags associated with content or feedback.

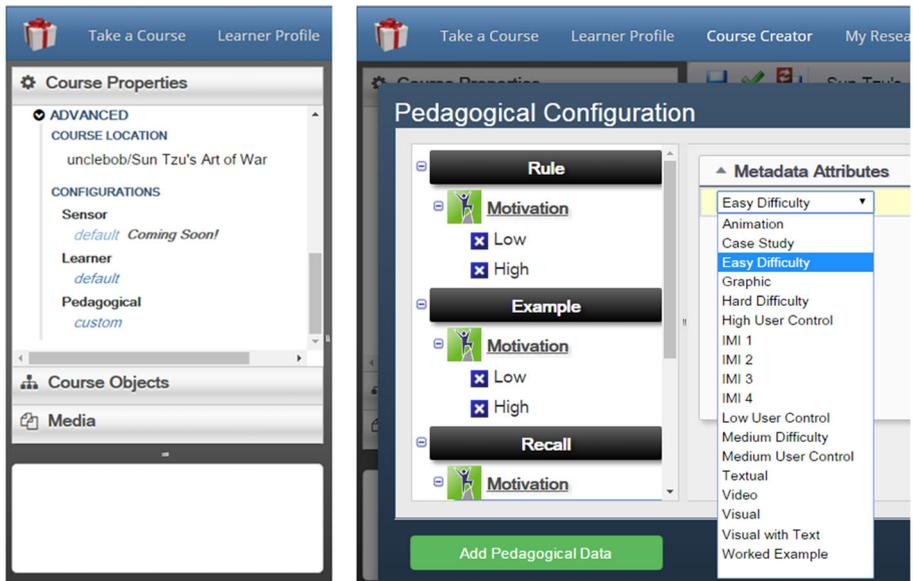


Fig. 6 Pedagogical configuration authoring in GIFT: course properties for pedagogical configuration may be customized by modifying or adding new pedagogical data (rules) to affiliate learner traits/states with recommended content or feedback attributes

Based on our team meta-analysis, we will need to develop a set of rules or agent policies that define similar relationships for teamwork strategies. Our goal is that assessment (classification of teamwork states) would be determined by the occurrences or frequency of the behavioral markers identified in our meta-analysis. This might be accomplished via semantic analysis of speech or text or sociometric badges (Pentland 2012; Calacci et al. 2016; Lederman et al. 2016). Other methods for classifying team behaviors and states of interest might include measures of voice levels and inflection of male and female team members (Titze 1989), content analysis of learner speech or text (Gottschalk and Gleser 1969) or understanding of non-verbal communications and behaviors and their influence on team states (Rosenberg and Sillince 2000; Kelly and Barsade 2001). It is likely that a combination of these methods will be needed to support accurate team state classification. Considerations in the design of classification methods include not only accuracy, but also speed to support real-time or near real-time feedback and unobtrusiveness to allow the team uninterrupted learning experiences.

A next step could be to associate the classification of any teamwork state (e.g., trust) with Johnson, Johnson & Stanne sentence openers to gradually train team members to communicate optimally in GIFT-mediated tutoring experiences.

Moving forward, as GIFT evolves from a decision-tree based architecture into a multi-agent architecture, software agents will be developed to detect and interpret individual learner behaviors as they relate to the team concepts identified herein. Our design goals for these agents are for them to be reactive, proactive, and cooperative. One implementation under consideration would develop a generalized personal agent assigned to detect and understand the behaviors of each individual on a team. While the focus of these agents would be on team members, they must also be cognizant of and responsive to changing conditions in the environment. There must be an understanding

within this agent architecture of how individual behaviors relate to team goals and team tutoring policies. Agents must be active in enforcing and updating policies to optimize team goals. The behavioral markers identified form a foundation for the measurement and assessment of team states. Policies based on these markers should drive actions by ITSs as they perceive and then select appropriate strategies to optimize team learning, performance, satisfaction, and viability.

Perception-action coupling (Fig. 7) highlights the need for the agent-based tutor to be cognizant of the changing conditions of both the environment and the learner. These *percepts* are used by the agent-based tutor to evaluate the continuous effectiveness of policies used to drive tutor actions. The evaluation of policies may lead to changes through reinforcement learning mechanisms and may affect a change to future tutor actions. The separate treatment of perception-action cycles for the learner and the environment has a theoretical basis in Vygotsky’s (1987), Zone of Proximal Development (ZPD). ZPD is the area within a learning experience (e.g., scenario or problem-solving event) where the challenge level of the experience is perfectly balanced with the learner’s competence (e.g., ability to solve the problem). According to Vygotsky, this balance is necessary to keep the learner engaged in the learning process.

For team tutoring to be effective, the agent-based tutor must have a model of each learner’s domain competency based on their assigned role(s) and task(s). The engagement of every team member can affect the learning, performance, satisfaction, and viability of the team. So it is important for the tutor to perceive interactions between each individual learner and the environment and understand the impact of those interactions on learning objectives. It is also critical to perceive and understand the impact of interactions between team members. All of these interactions constitute the basis for a team model as identified by the behavioral markers discussed herein.

In addition to being reactive to changes in the environment and learners on a team, agents should also be proactive in taking initiative to progress toward team goals. They should be capable of recognizing conditions which represent opportunities, and they should also be capable of learning and adapting to enhance the instructional

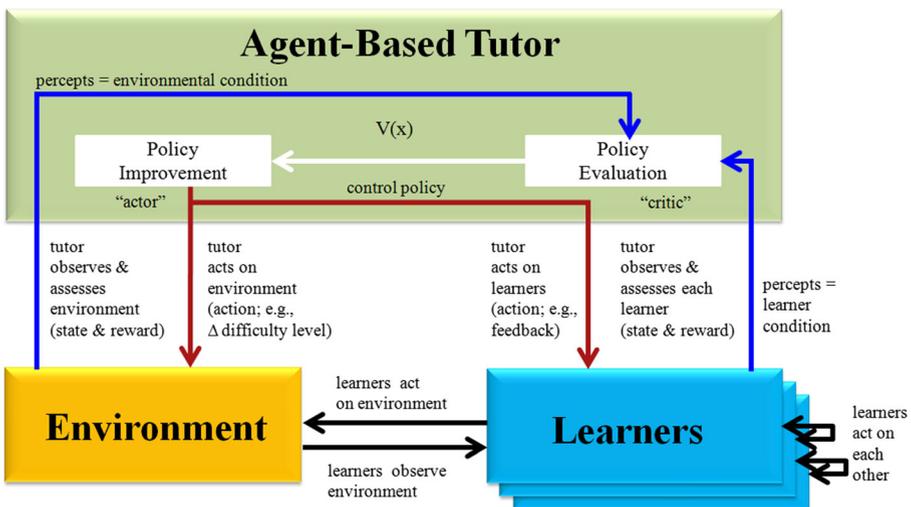


Fig. 7 Perception-action coupling in agent-based ITSs

experiences of individuals and teams. Finally, agents should be cooperative in sharing information about learners on a team to develop a comprehensive picture of the whole environment including the state(s) of the team. Together GIFT agents should work collaboratively to help learners achieve long-term learning goals.

Mechanisms have already been implemented in GIFT to track individual achievements during and between tutoring sessions. These achievement statements form the basis of a long-term learner model (LTLM) which is maintained in a learner record store. Mechanisms are needed to expand the LTLM to include team-based achievements, and to classify team competency states based on individual learner competencies. Finally, we must address how a GIFT-based ITS will respond effectively to the classification of various team states.

Discussion

While a significant focus of the research described in this article is on the significant advantages offered by a domain independent approach to team tutoring, we are compelled to also discuss potential limitations in domain independent approaches to ITS design. The flexibility in authoring and potential reuse offered by domain independent approaches may be offset by lower levels of effectiveness. In GIFT, we attempt to overcome this potential loss of effectiveness by balancing generalizable strategies (domain independent plans for action) with domain dependent tactics (domain dependent actions by the tutor).

In the application of our meta-analysis to GIFT, our goal was to identify significant antecedents to team learning and team performance. The structural equation models (SEMs) produced as part of this meta-analysis were derived by examining a variety of approaches and effect sizes in the literature. While all the effect sizes included in this study were relevant to team instruction, we caution ITS designers and authors that the studies spanned a variety of task domains. Generalizing the results across task domains and applying them to specific domains may or may not provide the same level of effectiveness as in the original domain dependent study included in our meta-analysis. In other words, conflict was negatively correlated to team performance and explained only 3% of its variance. Although identifying behavioral markers of conflict should aid the classification of a team state of conflict with a high level of accuracy, the resolution of conflict facilitated by the ITS may or may not explain the same level of performance variance. The application of SEMs produced in this study will not produce the same effect in every instance for every team.

Recommendations and Future Research

The focus of the meta-analysis described herein is to identify behavioral markers which might be used to classify behavioral states (e.g., trust or conflict) which significantly contribute to either the performance or learning of teams. A next step in the research of effective adaptive instruction for teams is to expand the current meta-analysis and then update the structural equation models (SEMs) for each of the seven primary themes (communication, cooperation, coordination, cognition, coaching, conflict, and

conditions) as antecedents of our outcomes of interest (i.e. learning, performance, satisfaction, and viability). This should enhance the confidence of the SEMs developed herein by providing additional power to confirm relationships between antecedents and our outcomes.

Specifically, the relationship between conflict management and team learning, psychological safety and team learning, cohesion and team learning, interpersonal processes and team performance, mutual support and team satisfaction, and all of the antecedents to team viability would benefit from the identification of additional studies to enhance confidence in the stated results. Future research may also include the analysis of these outcomes and new perhaps team outcomes (e.g. antecedents for transfer of training to new training or operational experiences). Over time, the collection of data related to the influence of these behaviors on team outcomes will refine our understanding and drive future agent-based policies for ITSs.

Low-cost, unobtrusive sensing methods should continue to pave the way to add real-time physiological classifiers and resulting markers to the current list of behavioral markers. Along with behavioral markers, physiological markers should provide confirmatory evidence of both individual learner and team states. The addition of physiological measures should increase the classification accuracy of team models and allow ITSs to apply more effective strategies for team performance and team learning.

While team mental models only accounted for 10% of the variance in team performance, their importance to conveying learning objectives and progress, maintaining engagement, and normalizing expectations should not be underestimated. Specifically, behavioral and physiological measures are needed to inform a team cognitive state model capable of assessing mental workload and engagement for each individual member and across the team as this informs the tutor's selection of team strategies and tactics within GIFT (Fletcher and Sottilare 2013).

Finally, referring back to our review of the related research in the AIED literature, we recommend moving forward with implementation of the following functional capabilities in GIFT's testbed function:

- Neuronal synchrony – evaluate low cost, portable neurophysiological sensors and develop interfaces/sensor configurations for GIFT to support additional research on the effect of neuronal synchrony in teams across task domains and populations
- Cooperative learning – implement Soller's (2001) Collaborative Learning Conversation Skill Taxonomy to support additional research on cooperative learning and to tie the behavioral markers and classification identified in this meta-analysis to concrete actions by the tutor to encourage improved teamwork skills.
- Cooperative learning – examine and cross-reference strategies in Adamson et al. (2014) academical productive talk (APT) strategies with generalized strategies resulting from this meta-analysis.
- Intelligent support for learning in groups – implement capabilities in GIFT to support evaluation of the effect of various team tutoring strategies across various task domains, team member roles, and leadership styles to determine optimal support strategies for intelligent agents.
- Remediation for negative teamwork behaviors - examine the effect of methods to remediate negative teamwork states.

- Broad application of teamwork models – apply the antecedents discussed herein to taskwork team training and collaborative learning.
- External system interface design – extend GIFT application program interface to interact with external team-based systems including team proxies (e.g., AutoTutor dialogues and Betty’s Brain) to enable team-based tutoring research.

Conclusions

The core contributions of this research provide: 1) evidence of the effect of teamwork behaviors in identifying antecedents of team performance and team learning; 2) structural equation models for the team performance and team learning for use in ITSs; and 3) identification of behavioral markers which can be used as measures for assessment of team performance and team learning during adaptive instruction. Each of the significant antecedents described herein form a model to support the adaptive tutoring of teams.

Collective efficacy, cohesion, communication, and leadership were significant antecedents of team performance. Trust, cohesion, conflict, and conflict management were identified as significant antecedents of team learning. While direct antecedents of team performance and team learning were identified, indirect influencers were also identified in our ontologies. Separate ontologies for team performance and team learning were constructed based on our MASEM process, but an expansion of this analysis over the next few years as new studies are conducted should result in even higher degrees of confidence for some of the meta-analytic results shown herein.

To provide a means to measure team outcomes, we identified six sets of behavioral markers: trust, collective efficacy, cohesion, communication, and conflict/conflict management. The markers formed the basis of a process to identify team member behaviors associated with antecedents of team performance and team learning. They also pave the way to identify methods to acquire team state data and classify team states. We also reviewed next steps for applying our meta-analytic findings to GIFT and discussed additional research and next steps needed to fully realize our vision for the adaptive tutoring of teams.

The learning and performance models discussed in this article are focused on the classification of teamwork behaviors. To this end, we have developed a set of markers that can be used within ITSs to identify various teamwork behaviors that indicate both individual and team learning and performance. Whereas the work of Adamson et al. (2014) was focused on the facilitation of academically productive talk (APT), our goal was to examine group interactions to determine if team communication was productive and supportive of learning and performance.

Our approach is sufficiently different from Adamson and others in the AIED and CSCL literature in that it focused on finding effect sizes in a large swath of the team literature to construct a relational model of teamwork. While Adamson and colleagues did not argue that the foundation provided in their approach was sufficient, they did stress their results to be part of “a larger, more thorough and systematic investigation of the space of possibilities” (Adamson et al. 2014, p. 121). We concur and also do not argue that our approach is a complete solution, but is a necessary step toward a needed much larger and complex solution.

Acknowledgements The research described herein has been sponsored by the U.S. Army Research Laboratory. The statements and opinions expressed in this article do not necessarily reflect the position or the policy of the United States Government, and no official endorsement should be inferred.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Adamson, D., Ashe, C., Jang, H., Yaron, D., & Rosé, C. (2013). Intensification of group knowledge exchange with academically productive talk agents. Proceedings of the 10th International Conference on Computer Supported Collaborative Learning, Madison Wisconsin, July 2013.
- Adamson, D., Dyke, G., Jang, H., & Rosé, C. P. (2014). Towards an agile approach to adapting dynamic collaboration support to student needs. *International Journal of Artificial Intelligence in Education*, 24(1), 92–124.
- Ai, H., Kumar, R., Nguyen, D., Nagasunder, A., & Rosé, C. P. (2010). Exploring the effectiveness of social capabilities and goal alignment in computer supported collaborative learning. In Proceedings of Intelligent Tutoring Systems, Lecture Notes in Computer Science, volume 6095, pp 134–143.
- Arthur Jr., W., Bennett Jr., W., & Huffcutt, A. I. (2001). *Conducting meta-analysis using SAS*. Mahwah: Erlbaum.
- Bandura, A. (1997). *Self-efficacy: The exercise of control*. London: Macmillan Publishers.
- Barker, J., Tjosvold, D., & Andrews, I. R. (1988). Conflict approaches of effective and ineffective project managers: a field study in a matrix organization. *Journal of Management Studies*, 25(2), 167–178.
- Beal, D. J., Cohen, R. R., Burke, M. J., & McLendon, C. L. (2003). Cohesion and performance in groups: a meta-analytic clarification of construct relations. *Journal of Applied Psychology*, 88(6), 989–1004.
- Bell, S. T. (2007). Deep-level composition variables as predictors of team performance: a meta-analysis. *Journal of Applied Psychology*, 92, 595–615.
- Biernat, M., & Manis, M. (1994). Shifting standards and stereotype-based judgments. *Journal of Personality and Social Psychology*, 66(1), 5–20.
- Biswas, G., Segedy, J. R., & Bunchongchit, K. (2016). From design to implementation to practice a learning by teaching system: Betty's Brain. *International Journal of Artificial Intelligence in Education*, 26(1), 350–364.
- Bloom, B. S. (1984). The 2 sigma problem: the search for methods of group instruction as effective as one-to-one tutoring. *Educational Researcher*, 13(6), 4–16.
- Bonner, D., Slavina, A., MacAllister, A., Holub, J., Gilbert, S., Sinatra, A. M., Winer, E., & Dorneich, M. (2016). The hidden challenges of team tutor development. In R. Sottolare & S. Ososky (Eds.), *Proceedings of 4th Annual GIFT Users Symposium (GIFTSym4)* (pp. 49–60). Orlando: Army Research Laboratory.
- Borman, W. C., & Smith, T. N. (2012). *The use of objective measures as criteria in I/O psychology. The Oxford handbook of personnel assessment and selection* (p. 532). New York: Oxford University Press, Inc.
- Bound, J. (1989). *Self-reported vs. objective measures of health in retirement models (No. w2997)*. Cambridge: National Bureau of Economic Research.
- Bray, S. R. (2004). Collective efficacy, group goals, and group performance of a muscular endurance task. *Small Group Research*, 35(2), 230–238.
- Budescu, D. V., & Bruderman, M. (1995). The relationship between the illusion of control and the desirability bias. *Journal of Behavioral Decision Making*, 8(2), 109–125.
- Bunderson, J. S., & Sutcliffe, K. M. (2002). Comparing alternative conceptualizations of functional diversity in management teams: process and performance effects. *Academy of Management Journal*, 45(5), 875–893.
- Burke, C. S., Stagl, K. C., Klein, C., Goodwin, G. F., Salas, E., & Halpin, S. M. (2006). What type of leadership behaviors are functional in teams? A meta-analysis. *The Leadership Quarterly*, 17, 288–307.

- Burns, H. L., & Capps, C. G. (2013). Foundations of intelligent tutoring systems: An introduction. In M. C. Polson & J. J. Richardson (Eds.), *Foundations of intelligent tutoring systems* (pp. 1–19). Hillsdale: Psychology Press.
- Byrne, M. (2001). Hermeneutics as a methodology for textual analysis. *AORN Journal*, 73(5), 968–970.
- Cai, Z., Feng, S., Baer, W., & Graesser, A. (2014). Instructional strategies in Trialogue-based intelligent tutoring systems. *Design Recommendations for Intelligent Tutoring Systems*, 2, 225–235.
- Calacci, D., Lederman, O., Shrier, D., & Pentland, A. S. (2016). Breakout: An open measurement and intervention tool for distributed peer learning groups. Paper presented at the International Conference on Social Computing, Behavioral-Cultural Modeling, & Prediction and Behavior Representation in Modeling and Simulation (SBP-BRiMS), Washington, D.C.
- Campion, M. A., Medsker, G. J., & Higgs, A. C. (1993). Relations between work group characteristics and effectiveness: implications for designing effective work groups. *Personnel Psychology*, 46(4), 823–847.
- Cannon-Bowers, J. A., & Bowers, C. (2011). Team development and functioning. In S. Zedeck (Ed.), *APA handbook of industrial and organizational psychology, Vol 1: Building and developing the organization* (pp. 597–650). Washington, DC: American Psychological Association.
- Cannon-Bowers, J. A., Tannenbaum, S. I., Salas, E., & Volpe, C. E. (1995). *Defining competencies and establishing team training requirements. Team effectiveness and decision making in organizations* (pp. 333–380). San Francisco: Jossey-Bass.
- Carless, S. A., & De Paola, C. (2000). The measurement of cohesion in work teams. *Small Group Research*, 31(1), 71–88.
- Carron, A. V., Widmeyer, W. N., & Brawley, L. R. (1985). The development of an instrument to assess cohesion in sport teams: the group environment questionnaire. *Journal of Sport Psychology*, 7(3), 244–266.
- Chang, A., & Bordia, P. (2001). A multidimensional approach to the group cohesion-group performance relationship. *Small Group Research*, 32(4), 379–405.
- Chang, J., Sy, T., & Choi, J. (2012). Team emotional intelligence and performance: interactive dynamics between leaders and members. *Small Group Research*, 43, 75–104. doi:10.1177/1046496411415692.
- Chaudhuri, S., Kumar, R., Howley, I., & Rosé, C. P. (2009). Engaging collaborative learners with helping agents. In Proceedings of the 2009 conference on artificial intelligence in education: Building learning systems that care: From knowledge representation to affective modeling (pp. 365–372). IOS Press.
- Chen, G., Thomas, B., & Wallace, J. C. (2005). A multilevel examination of the relationships among training outcomes, mediating regulatory processes, and adaptive performance. *Journal of Applied Psychology*, 90(5), 827.
- Cheng, H. H., & Yang, H. L. (2011). Student team projects in information systems development: measuring collective creative efficacy. *Australasian Journal of Educational Technology*, 27(6), 881–895.
- Cook, J., & Wall, T. (1980). New work attitude measures of trust, organizational commitment and personal need non-fulfillment. *Journal of Occupational Psychology*, 53(1), 39–52.
- Costa, A. C. (2003). *Understanding the nature and the antecedents of trust within work teams. The trust process in organizations: Empirical studies of the determinants and the process of trust development* (pp. 105–124). Cheltenham: Edward Elgar Publishing
- Craig, T. Y., & Kelly, J. R. (1999). Group cohesiveness and creative performance. *Group Dynamics: Theory, research, and practice*, 3(4), 243.
- Cronin, M. A., Bezrukova, K., Weingart, L. R., & Tinsley, C. H. (2011). Subgroups within a team: the role of cognitive and affective integration. *Journal of Organizational Behavior*, 32(6), 831–849.
- Davis, J. B. (2007). *Statistics using SAS enterprise guide*. Cary: SAS Institute.
- De Dreu, C. K. (2007). Cooperative outcome interdependence, task reflexivity, and team effectiveness: a motivated information processing perspective. *Journal of Applied Psychology*, 92(3), 628.
- De Dreu, C. K., & Weingart, L. R. (2003). Task versus relationship conflict, team performance, and team member satisfaction: a meta-analysis. *Journal of Applied Psychology*, 88(4), 741–749.
- De Jong, B. A., & Elfring, T. (2010). How does trust affect the performance of ongoing teams? The mediating role of reflexivity, monitoring, and effort. *Academy of Management Journal*, 53(3), 535–549.
- Dennis, M., Masthoff, J., & Mellish, C. (2016). Adapting progress feedback and emotional support to learner personality. *International Journal of Artificial Intelligence in Education*, 26(3), 877–931.
- Dillenbourg, P. (1999). What do you mean by collaborative learning. *Collaborative-learning: Cognitive and Computational Approaches*, 1, 1–15.
- Dillenbourg, P., & Hong, F. (2008). The mechanics of CSCL macro scripts. *The International Journal of Computer-Supported Collaborative Learning*, 3(1), 5–23.
- Dirks, K. T. (2000). Trust in leadership and team performance: evidence from NCAA basketball. *Journal of Applied Psychology*, 85(6), 1004.

- Dorsey, D., Russell, S., Keil, C., Campbell, G., Van Buskirk, W., & Schuck, P. (2009). *Measuring teams in action: Automated performance measurement and feedback in simulation-based training. Team effectiveness in complex organizations: Cross-disciplinary perspectives and approaches* (pp. 351–381). Routledge: Taylor and Francis.
- du Boulay, B. (2016). Recent meta-reviews and meta-analyses of AIED systems. *International Journal of Artificial Intelligence in Education*, 1(26), 536–537.
- Dyer, L. (1984). Studying human resource strategy: an approach and an agenda. *Industrial Relations: A Journal of Economy and Society*, 23(2), 156–169.
- Dyke, G., Adamson, D., Howley, I., & Rosé, C. P. (2013). Enhancing scientific reasoning and explanation skills with conversational agents. *IEEE Transactions on Learning Technologies*, 6(3), 240–247.
- Edmonds, W. A., Tenenbaum, G., Kamata, A., & Johnson, M. B. (2009). The role of collective efficacy in adventure racing teams. *Small Group Research*, 40(2), 163–180.
- Edmondson, A., Bohmer, R. M., & Pisano, G. P. (2001). Disrupted routines: Team learning and new technology implementation in hospitals. *Administrative Science Quarterly*, 46(4), 685–716.
- Erkens, G., & Janssen, J. (2008). Automatic coding of dialogue acts in collaboration protocols. *International Journal of Computer Supported Collaborative Learning*, 3, 447–470.
- Espinosa, J. A., Cummings, J. N., & Pickering, C. (2012). Time separation, coordination, and performance in technical teams. *IEEE Transactions on Engineering Management*, 59(1), 91–103.
- Faraj, S., & Sproull, L. (2000). Coordinating expertise in software development teams. *Management Science*, 46(12), 1554–1568.
- Fletcher, T. D., & Major, D. A. (2006). The effects of communication modality on performance and self-ratings of teamwork components. *Journal of Computer-Mediated Communication*, 11(2), 557–576.
- Fletcher, J. D., & Sottolare, R. (2013). Shared mental models of cognition for intelligent tutoring of teams. In R. Sottolare, A. Graesser, X. Hu, & H. Holden (Eds.), *Design recommendations for intelligent tutoring systems: Volume 1- learner modeling*. Orlando: Army Research Laboratory ISBN 978-0-9893923-0-3.
- Gajendran, R. S., & Joshi, A. (2012). Innovation in globally distributed teams: the role of LMX, communication frequency, and member influence on team decisions. *Journal of Applied Psychology*, 97(6), 1252.
- Goldberg, B., Brawner, K., Sottolare, R., Tarr, R., Billings, D., & Malone, M. (2012). Use of evidence-based strategies to expand extensibility of adaptive tutoring technologies. In Proceedings of the *Interservice/Industry Training Simulation & Education Conference*, Orlando, Florida, December 2012.
- Gottschalk, L. A., & Gleser, G. C. (1969). *The measurement of psychological states through the content analysis of verbal behavior*. Oakland: University of California Press.
- Greer, J. E., Mccalla, G., Collins, J. A., Kumar, V. S., Meagher, P., & Vassileva, J. (1998). Supporting peer help and collaboration in distributed workplace environments. *International Journal of Artificial Intelligence in Education (IJAIED)*, 9, 159–177.
- Greer, L. L., Homan, A. C., De Hoogh, A. H., & Den Hartog, D. N. (2012). Tainted visions: the effect of visionary leader behaviors and leader categorization tendencies on the financial performance of ethnically diverse teams. *Journal of Applied Psychology*, 97(1), 203.
- Gupta, A., Mattarelli, E., Seshasai, S., & Broschak, J. (2009). Use of collaborative technologies and knowledge sharing in co-located and distributed teams: towards the 24-h knowledge factory. *The Journal of Strategic Information Systems*, 18(3), 147–161.
- Gupta, V. K., Huang, R., & Niranjan, S. (2010). A longitudinal examination of the relationship between team leadership and performance. *Journal of Leadership & Organizational Studies*, 17(4), 335–350.
- Guzzo, R. A., Yost, P. R., Campbell, R. J., & Shea, G. P. (1993). Potency in groups: articulating a construct. *British Journal of Social Psychology*, 32(1), 87–106.
- Haas, M. R. (2006). Acquiring and applying knowledge in transnational teams: The roles of cosmopolitans and locals. *Organization Science*, 17(3), 367–384.
- Hackman, J. R. (1987). The design of work teams. In J. W. Lorsch (Ed.), *Handbook of organizational behavior* (pp. 315–342). Englewood Cliffs: Prentice-Hall.
- Hackman, J. R., & Wageman, R. (2005). A theory of team coaching. *Academy of Management Review*, 30(2), 269–287.
- Hall, L., Tazzyman, S., Hume, C., Endrass, B., Lim, M. Y., Hofstede, G., et al. (2015). Learning to overcome cultural conflict through engaging with intelligent agents in synthetic cultures. *International Journal of Artificial Intelligence in Education*, 25(2), 291–317.
- Henry, K. B., Arrow, H., & Carini, B. (1999). A tripartite model of group identification theory and measurement. *Small Group Research*, 30(5), 558–581.
- Hirst, G. (2009). Effects of membership change on open discussion and team performance: the moderating role of team tenure. *European Journal of Work and Organizational Psychology*, 18(2), 231–249.

- Hobman, E. V., & Bordia, P. (2006). The role of team identification in the dissimilarity-conflict relationship. *Group Processes & Intergroup Relations*, 9(4), 483–507.
- Hoegl, M., & Gemuenden, H. G. (2001). Teamwork quality and the success of innovative projects: a theoretical concept and empirical evidence. *Organization Science*, 12(4), 435–449.
- Hsu, M. H., Chen, I. Y. L., Chiu, C. M., & Ju, T. L. (2007). Exploring the antecedents of team performance in collaborative learning of computer software. *Computers & Education*, 48(4), 700–718.
- Hu, L. T., & Bentler, P. M. (1995). Evaluating model fit. In R. Hoyle (Ed.), *Structural equation modeling: Issues, concepts, and applications* (pp. 76–99). Newbury Park: Sage.
- Huang, J. C. (2010). Unbundling task conflict and relationship conflict: the moderating role of team goal orientation and conflict management. *International Journal of Conflict Management*, 21(3), 334–355.
- Huang, S., & Cummings, J. N. (2011). When critical knowledge is most critical: centralization in knowledge-intensive teams. *Small Group Research*, 42(6), 669–699.
- Hunter, J. E., & Schmidt, F. L. (2004). *Methods of meta-analysis: Correcting error and bias in research findings*. London: Sage.
- Israel, J., & Aiken, R. (2007). Supporting collaborative learning with an intelligent web-based system. *International Journal of Artificial Intelligence in Education*, 17(1), 3–40.
- Jehn, K. A. (1995). A multimethod examination of the benefits and detriments of intragroup conflict. *Administrative Science Quarterly*, 40(2), 256–282.
- Jehn, K. A., & Mannix, E. A. (2001). The dynamic nature of conflict: a longitudinal study of intragroup conflict and group performance. *Academy of Management Journal*, 44(2), 238–251.
- Jehn, K. A., Chadwick, C., & Thatcher, S. M. (1997). To agree or not to agree: the effects of value congruence, individual demographic dissimilarity, and conflict on workgroup outcomes. *International Journal of Conflict Management*, 8(4), 287–305.
- Johns, J. H., Bickel, M. D., Blades, A. C., Creel, J. B., Gatling, W. S., Hinkle, J. M., et al. (1984). *Cohesion in the US military*. Washington DC: National Defense University.
- Johnson, R. T., & Johnson, D. W. (1986). Cooperative learning in the science classroom. *Science and Children*, 24, 31–32.
- Johnson, D. W., & Johnson, R. T. (1999). What makes cooperative learning work. In D. Kluge, S. McGuire, D. Johnson, & R. Johnson (Eds), *JALT applied materials: cooperative learning* (pp. 23–36). Tokyo: Japan Association for Language Teaching.
- Johnson, D. W., Johnson, R. T., & Stanne, M. B. (2000). *Cooperative learning methods: A meta-analysis*. Minneapolis: University of Minnesota.
- Jones, G. R. (1986). Socialization tactics, self-efficacy, and newcomers' adjustments to organizations. *Academy of Management Journal*, 29(2), 262–279.
- Jong, A. D., Ruyter, K. D., & Lemmink, J. (2005). Service climate in self-managing teams: mapping the linkage of team member perceptions and service performance outcomes in a business-to-business setting. *Journal of Management Studies*, 42(8), 1593–1620.
- Jöreskog, K. (1969). A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika*, 34, 183–202.
- Jöreskog, K. G., & Sörbom, D. (2004). *LISREL 8.71: Structural equation modeling with the SIMPLIS command language*. Chicago: Scientific Software International.
- Kelly, J. R., & Barsade, S. G. (2001). Mood and emotions in small groups and work teams. *Organizational Behavior and Human Decision Processes*, 86(1), 99–130.
- Klein, C., DiazGranados, D., Salas, E., Le, H., Burke, C. S., Lyons, R., & Goodwin, G. F. (2009). Does team building work? *Small Group Research*, 40, 181–222.
- Kostopoulos, K. C., & Bozionelos, N. (2011). Team exploratory and exploitative learning: psychological safety, task conflict, and team performance. *Group & Organization Management*, 36(3), 385–415.
- Kulik, J. A., & Fletcher, J. D. (2015). Effectiveness of Intelligent Tutoring Systems: A Meta-Analytic Review. *Review of Educational Research*, X(X), 1–37. doi:10.3102/0034654315581420.
- Kumar, R., & Rose, C. P. (2011). Architecture for building conversational agents that support collaborative learning. *IEEE Transactions on Learning Technologies*, 4(1), 21–34.
- Kumar, R., & Kim, J. (2014). Special Issue on Intelligent Support for Learning in Groups. *International Journal of Artificial Intelligence in Education*, 24, 1. doi:10.1007/s40593-013-0013-5.
- Kumar, R., Rosé, C. P., Wang, Y. C., Joshi, M., & Robinson, A. (2007). Tutorial dialogue as adaptive collaborative learning support. Proceedings of the 2007 conference on artificial intelligence in education, 383–390.
- Kumar, R., Ai, H., Beuth, J., & Rosé, C. P. (2010). Socially-capable conversational tutors can be effective in collaborative learning situations. In Proceedings of intelligent tutoring systems, lecture notes in computer science volume 6095, pp 156–164.

- Lederman, O., Calacci, D., MacMullen, A., Fehder, D. C., Murray, F. E., & Pentland, A. S. (2016). Open badges: A low-cost toolkit for measuring team communication and dynamics. Paper presented at the International Conference on Social Computing, Behavioral-Cultural Modeling, & Prediction and Behavior Representation in Modeling and Simulation (SBP-BRiMS), Washington, D.C.
- Lee, P., Gillespie, N., Mann, L., & Wearing, A. (2010). Leadership and trust: Their effect on knowledge sharing and team performance. *Management learning*, *41*(4), 473–491.
- Leelawong, K., & Biswas, G. (2008). Designing learning by teaching agents: the Betty's Brain system. *International Journal of Artificial Intelligence in Education*, *18*(3), 181–208.
- Lehman, B., D'Mello, S., Strain, A., Mills, C., Gross, M., Dobbins, A., et al. (2013). Inducing and tracking confusion with contradictions during complex learning. *International Journal of Artificial Intelligence in Education*, *22*(1–2), 85–105.
- Lent, R. W., Schmidt, J., & Schmidt, L. (2006). Collective efficacy beliefs in student work teams: relation to self-efficacy, cohesion, and performance. *Journal of Vocational Behavior*, *68*(1), 73–84.
- Li, H., Rosenthal, R., & Rubin, D. B. (1996). Reliability of measurement in psychology: from Spearman-Brown to maximal reliability. *Psychological Methods*, *1*(1), 98.
- Luhtanen, R., & Crocker, J. (1992). A collective self-esteem scale: self-evaluation of one's social identity. *Personality and Social Psychology Bulletin*, *18*(3), 302–318.
- Ma, W., Adesope, O. O., Nesbit, J. C., & Liu, Q. (2014). Intelligent tutoring systems and learning outcomes: a meta-analysis. *Journal of Educational Psychology*, *106*(4), 901–918. doi:10.1037/a0037123.
- Mathieu, J. E., Rapp, T. L., Maynard, M. T., & Mangos, P. M. (2009). Interactive effects of team and task shared mental models as related to air traffic controllers' collective efficacy and effectiveness. *Human Performance*, *23*(1), 22–40.
- Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organizational trust. *Academy of Management Review*, *20*, 709–734.
- McCroskey, J. C., & McCain, T. A. (1974). *The measurement of interpersonal attraction*. Paper presented at the Annual Convention of the Western Speech Communication Assn. Honolulu, HI: November 1972 (pp. 261–266)
- McManus, M., & Aiken, R. (1993). The group leader paradigm in an intelligent collaborative learning system. Proceedings of the 7th world conference on Artificial Intelligence in Education (AIED 93), Edinburgh, Scotland, 249–256.
- McManus, M. M., & Aiken, R. M. (1995). Monitoring computer-based collaborative problem solving. *Journal of Artificial Intelligence in Education*, *6*(4), 307–336.
- McManus, M. M., & Aiken, R. M. (2016). Supporting effective collaboration: Using a rearview mirror to look forward. *International Journal of Artificial Intelligence in Education*, *26*(1), 365–377.
- Meister, D. (1985). *Behavioral analysis and measurement methods*. Hoboken: Wiley-Interscience.
- Merrill, D. (2015). Component display theory. Instructional Design.org. Retrieved from <http://www.instructionaldesign.org/theories/component-display.html>
- Mesmer-Magnus, J., & Viswesvaran, C. (2010). The role of pre-training interventions in learning: a meta-analysis and integrative review. *Human Resource Management Review*, *20*(4), 261–282.
- Miller, D. C. (1964). *Handbook of research design and social measurement*. New York: David McKay Company.
- Muehlenbrock, M., Tewissen, F., & Hoppe, U. (1998). A framework system for intelligent support in open distributed learning environments. *International Journal of Artificial Intelligence in Education (IJAIED)*, *9*, 256–274.
- North Atlantic Treaty Organization. (2012). *Technical area plan for the human factors and medicine Panel's research task group 237, Assessment of Intelligent Tutoring System Technologies and Opportunities*.
- Nunnally, J. C. (1978). *Psychometric theory*. New York: McGraw-Hill.
- O'Reilly III, C. A., Caldwell, D. F., & Barnett, W. P. (1989). Work group demography, social integration, and turnover. *Administrative Science Quarterly*, *34*, 21–37.
- Pane, J. F., Griffin, B. A., McCaffrey, D. F., & Karam, R. (2014). Effectiveness of cognitive tutor algebra I at scale. *Educational Evaluation and Policy Analysis*, *36*(2), 127–144. doi:10.3102/0162373713507480.
- Pearson, A. W., Ensley, M. D., & Amason, A. C. (2002). An assessment and refinement of Jehn's intragroup conflict scale. *International Journal of Conflict Management*, *13*(2), 110–126.
- Pentland, A. (2012). The new science of building great teams. *Harvard Business Review*, *90*(4), 60–69.
- Podsakoff, P. M., MacKenzie, S. B., & Ahearne, M. (1997). Moderating effects of goal acceptance on the relationship between group cohesiveness and productivity. *Journal of Applied Psychology*, *82*(6), 974.
- Ponterotto, J. G., Gretchen, D., Utsey, S. O., Stracuzzi, T., & Saya, R. (2003). The multigroup ethnic identity measure (MEIM): psychometric review and further validity testing. *Educational and Psychological Measurement*, *63*(3), 502–515.

- Quintana, S. M., & Maxwell, S. E. (1999). Implications of recent developments in structural equation modeling for counseling psychology. *The Counseling Psychologist*, 27(4), 485–527.
- Riggs, M. L., & Knight, P. A. (1994). The impact of perceived group success-failure on motivational beliefs and attitudes: a causal model. *Journal of Applied Psychology*, 79(5), 755.
- Rispens, S. (2012). The influence of conflict issue importance on the co-occurrence of task and relationship conflict in teams. *Applied Psychology*, 61(3), 349–367.
- Rosé, C., & VanLehn, K. (2005). An evaluation of a hybrid language understanding approach for robust selection of tutoring goals. *International Journal of Artificial Intelligence in Education*, 15(4), 325–355.
- Rosé, C. P., Jordan, P., Ringenberg, M., Siler, S., VanLehn, K., & Weinstein, A. (2001). Interactive conceptual tutoring in Atlas-Andes. In Proceedings of AI in education 2001 conference, 151–153.
- Rosen, Y. (2015). Computer-based assessment of collaborative problem solving: exploring the feasibility of human-to-agent approach. *International Journal of Artificial Intelligence in Education*, 25(3), 380–406.
- Rosenberg, D., & Sillince, J. A. (2000). Verbal and nonverbal communication in computer mediated settings. *International Journal of Artificial Intelligence in Education (IJAIED)*, 11, 299–319.
- Rosenfeld, L. B., & Gilbert, J. R. (1989). The measurement of cohesion and its relationship to dimensions of self-disclosure in classroom settings. *Small Group Research*, 20(3), 291–301.
- Rutkowski, A. F., Saunders, C., Vogel, D., & Van Genuchten, M. (2007). “Is it already 4 am in your time zone?” focus immersion and temporal dissociation in virtual teams. *Small Group Research*, 38(1), 98–129.
- Salas, E. (2015). *Team training essentials: A research-based guide*. London: Routledge.
- Salas, E., Dickinson, T. L., Converse, S. A., & Tannenbaum, S. I. (1992). Toward an understanding of team performance and training. In R. W. Swezey, E. Salas, R. W. Swezey, & E. Salas (Eds.), *Teams: Their training and performance* (pp. 3–29). Westport: Ablex Publishing.
- Sargent, L. D., & Sue-Chan, C. (2001). Does diversity affect group efficacy? The intervening role of cohesion and task interdependence. *Small Group Research*, 32(4), 426–450.
- Schippers, M. C., Den Hartog, D. N., & Koopman, P. L. (2007). Reflexivity in teams: a measure and correlates. *Applied Psychology*, 56(2), 189–211.
- Schmidt, F. L., & Hunter, J. E. (1977). Development of a general solution to the problem of validity generalization. *Journal of Applied Psychology*, 62(5), 529.
- Shin, Y., & Choi, J. N. (2010). What makes a group of good citizens? The role of perceived group-level fit and critical psychological states in organizational teams. *Journal of Occupational and Organizational Psychology*, 83(2), 531–552.
- Shivers-Blackwell, S. L. (2004). Using role theory to examine determinants of transformational and transactional leader behavior. *Journal of Leadership & Organizational Studies*, 10(3), 41–50.
- Siebold, G. L. (1999). The evolution of the measurement of cohesion. *Military Psychology*, 11(1), 5.
- Singley, M. K., Fairweather, P. G. & Swerling, S. (1999). Team tutoring systems: Reifying roles in problem solving. *Proceedings of Computer-Support for Collaborative Learning (CSCL'99)*, Stanford, California, 538-548.
- Smith-Jentsch, K. A., Zeisig, R. L., Acton, B., & McPherson, J. A. (1998). Team dimensional training: A strategy for guided team self-correction. In J. A. Cannon-Bowers & E. Salas (Eds.), *Making decisions under stress: Implications for individual and team training* (pp. 271–297). Washington, DC: American Psychological Association.
- Smith-Jentsch, K. A., Cannon-Bowers, J. A., Tannenbaum, S. I., & Salas, E. (2008). Guided team self-correction: Impacts on team mental models, processes, and effectiveness. *Small Group Research*, 39(3), 303–327.
- Solansky, S. T. (2011). Team identification: a determining factor of performance. *Journal of Managerial Psychology*, 26(3), 235–246.
- Soller, A. (2001). Supporting social interaction in an intelligent collaborative learning system. *International Journal of Artificial Intelligence in Education (IJAIED)*, 12, 40–62.
- Sottolare, R. (2012). Considerations in the development of an ontology for a generalized intelligent framework for tutoring. *International Defense & Homeland Security Simulation Workshop in Proceedings of the I3M Conference*. Vienna, Austria, September 2012.
- Sottolare, R.A. & Holden, H.K. (2013). Motivations for a generalized intelligent framework for tutoring (GIFT) for authoring, instruction and analysis. In *Proceedings of the 1st Generalized Intelligent Framework for Tutoring (GIFT) Users Symposium at the Artificial Intelligence in Education (AIED) Conference 2013*, Memphis, Tennessee, June 2013.
- Sottolare, R. A., & Proctor, M. D. (2012). Passively classifying student mood and performance within intelligent tutors. *Educational Technology & Society*, 15(2), 101–114.

- Sottolare, R. A., Holden, H. K., Brawner, K. W., & Goldberg, B. S. (2011). Challenges and emerging concepts in the development of adaptive, computer-based tutoring systems for team training. *Interservice/Industry Training Systems & Education Conference*, Orlando, Florida, December 2011.
- Sottolare, R.A., Brawner, K.W., Goldberg, B.S. & Holden, H.K. (2012). The generalized intelligent framework for tutoring (GIFT). Concept paper released as part of GIFT software documentation. Orlando, FL: U.S. Army Research Laboratory – Human Research & Engineering Directorate (ARL-HRED). Retrieved from: https://gifttutoring.org/attachments/152/GIFTDescription_0.pdf
- Sottolare R., Ragusa C., Hoffman, M. & Goldberg B. (2013). Characterizing an adaptive tutoring learning effect chain for individual and team tutoring. In *Proceedings of the Interservice/Industry Training Simulation and Education Conference*; 2013 Dec; Orlando, FL.
- Steenbergen-Hu, S., & Cooper, H. (2013). A meta-analysis of the effectiveness of intelligent tutoring systems on K–12 students' mathematical learning. *Journal of Educational Psychology*, 105(4), 970–987. doi:10.1037/a0032447.
- Steenbergen-Hu, S., & Cooper, H. (2014). A meta-analysis of the effectiveness of intelligent tutoring systems on college students' academic learning. *Journal of Educational Psychology*, 106(2), 331–347. doi:10.1037/a0034752.
- Stevens, R., Berka, C., & Sprang, M. (2009a, October). Neurophysiologic collaboration patterns during team problem solving. In *Proceedings of the human factors and ergonomics Society annual meeting* (Vol. 53, no. 12, pp. 804–808). Thousand Oaks: SAGE Publications.
- Stevens, R. H., Galloway, T., Berka, C., & Sprang, M. (2009b, July). Can neurophysiologic synchronies provide a platform for adapting team performance?. In *International Conference on Foundations of Augmented Cognition* (pp. 658–667). Springer Berlin Heidelberg.
- Stevens, R., Gorman, J. C., Amazeen, P., Likens, A., & Galloway, T. (2013). The organizational neurodynamics of teams. *Nonlinear Dynamics, Psychology, and Life Sciences*, 17(1), 67–86.
- Tasa, K., Taggar, S., & Seijts, G. H. (2007). The development of collective efficacy in teams: a multilevel and longitudinal perspective. *Journal of Applied Psychology*, 92(1), 17.
- Tedesco, P. A. (2003). MARCo: building an artificial conflict mediator to support group planning interactions. *International Journal of Artificial Intelligence in Education*, 13(1), 117–155.
- Tegos, S., Demetriadis, S., & Tsiatsos, T. (2014). A configurable conversational agent to trigger students' productive dialogue: a pilot study in the CALL domain. *International Journal of Artificial Intelligence in Education*, 24(1), 62–91.
- Titze, I. R. (1989). Physiologic and acoustic differences between male and female voices. *The Journal of the Acoustical Society of America*, 85(4), 1699–1707.
- Tung, H. L., & Chang, Y. H. (2011). Effects of empowering leadership on performance in management team: Mediating effects of knowledge sharing and team cohesion. *Journal of Chinese Human Resources Management*, 2(1), 43–60.
- Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: heuristics and biases. *Science*, 185(4157), 1124–1131.
- U.S. Army Training, & Doctrine Command (TRADOC). (2011). *525–8-2 the US Army learning concept for 2015*. Washington, DC: Department of the Army.
- Van Berio, M. P. (1997). Team training vs. team building and cooperative learning: Defining the field of research (team training vs team building en cooperatief leren: Afbakening van het onderzoeksterrein) (No. TNO-TM-97-B019). HUMAN FACTORS RESEARCH INST TNO SOESTERBERG (NETHERLANDS).
- Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*, 3(1), 4–70.
- VanLehn, K. (2011). The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. *Educational Psychologist*, 46(4), 197–221.
- Viswesvaran, C., & Ones, D. S. (1995). Theory testing: combining psychometric meta-analysis and structural equations modeling. *Personnel Psychology*, 48(4), 865–885.
- Vygotsky, L. S., & Rieber, R. W. (1987). *The collected works of LS Vygotsky: Volume 1: Problems of general psychology, including the volume Thinking and Speech (Vol. 1)*. Berlin/Heidelberg: Springer Science and Business Media.
- Walker, E., & Ogan, A. (2016). We're in this together: Intentional Design of Social Relationships with AIED systems. *International Journal of Artificial Intelligence in Education*, 26(2), 713–729.
- Walker, E., Rummel, N., & Koedinger, K. R. (2014). Adaptive intelligent support to improve peer tutoring in algebra. *International Journal of Artificial Intelligence in Education*, 24(1), 33–61.

- Walther, J. B., & Bunz, U. (2005). The rules of virtual groups: trust, liking, and performance in computer-mediated communication. *Journal of Communication*, 55(4), 828–846.
- Walton, J., Dorneich, M. C., Gilbert, S., Bonner, D., Winer, E., & Ray, C. (2014). Modality and timing of team feedback: Implications for GIFT *Generalized Intelligent Framework for Tutoring (GIFT) Users Symposium (GIFTSym2)* (pp. 199–207).
- Watson, W. E., Michaelsen, L. K., & Sharp, W. (1991). Member competence, group interaction, and group decision making: a longitudinal study. *Journal of Applied Psychology*, 76(6), 803.
- Wiese, C. W., Shuffler, M. L., & Salas, E. (2015). Teamwork and team performance measurement. In J. Wright (Ed.), *International encyclopedia of the social & behavioral sciences* (pp. 96–103). Oxford: Pergamon.
- Woehr, D. J., Arciniega, L. M., & Poling, T. L. (2013). Exploring the effects of value diversity on team effectiveness. *Journal of Business and Psychology*, 28(1), 107–121.
- Wong, S. S. (2004). Distal and local group learning: performance trade-offs and tensions. *Organization Science*, 15(6), 645–656.
- Wong, A., Tjosvold, D., & Liu, C. (2009). Cross-functional team organizational citizenship behavior in China: shared vision and goal interdependence among departments. *Journal of Applied Social Psychology*, 39(12), 2879–2909.
- Yoo, J., & Kim, J. (2014). Can online discussion participation predict group project performance? Investigating the roles of linguistic features and participation patterns. *International Journal of Artificial Intelligence in Education*, 24(1), 8–32. doi:10.1007/s40593-013-0010-8.
- Zaccaro, S. J. (1991). Nonequivalent associations between forms of cohesiveness and group-related outcomes: evidence for multidimensionality. *The Journal of Social Psychology*, 131(3), 387–399.
- Zachary, W., Cannon-Bowers, J. A., Bilazarian, P., Kreckler, D. K., Lardieri, P. J., & Burns, J. (1999). The advanced embedded training system (AETS): an intelligent embedded tutoring system for tactical team training. *International Journal of Artificial Intelligence in Education (IJAIED)*, 10, 257–277.